Interactive comment on “A regime view of future atmospheric circulation changes in Northern mid-latitudes” by Federico Fabiano et al.
Anonymous Referee #1

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 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-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 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, a comment on the fact that the links may be due, at least partly, to a reversed causal relationship or to an external forcing influencing both processes in a similar way, are included to the discussion in the revised manuscript at lines 392-395: "*This analysis aims at finding significant relationships between the set of potential drivers and the WR frequency trends. Of course, a significant correlation between two quantities does not demonstrate the existence of a causal link - as they might be responding to a common external forcing - nor gives indication on the direction of such link. Nevertheless, this can provide an insight about the interconnection between mid-latitude climate variability and large scale changes in GCMs that deserve further investigation.*"

The definition 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 (TNAW) and the subpolar gyre/warming hole region (SPGW). We included a new comment in the discussion at lines 430-431, suggesting that it may also be that both the TNAW and the change in the Pacific WRs frequency share a common external driver: "*However, this significant regression may also be produced by a common external driver, influencing both the TNAW and the PNA regimes, like a change in the ENSO forcing (Fredriksen et al., 2020).*"

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, Hassanzadeh 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, Hassanzadeh 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 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 references to these works in the introduction at lines 79-82: *"Many works in literature studied how WRs are reproduced by GCMs, but mostly focused on the model performance in control or historical simulations (Dawson et al., 2012; Cattiaux et al., 2013b; Dawson and Palmer, 2015; Weisheimer et al., 2014; Strommen et al., 2019; Fabiano et al., 2020). Changes of WRs in CMIP5 projections were analyzed by Cattiaux et al. (2013a) and by Ullmann et al. (2014) for the EAT sector."*
We also added a comparison of our CMIP5 results with the Cattiaux ones in the discussion section at lines 326-328: "*The change of WR frequency observed for RCP8.5 is consistent with the result by Cattiaux et al. (2013a), although they observe a larger increase of NAO-, which may be due to a different treatment of the climatological mean state.*"

• 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)?

- A new figure (S11) was added to the supplementary material and is reported below. The figure (left panel) 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 right panel).

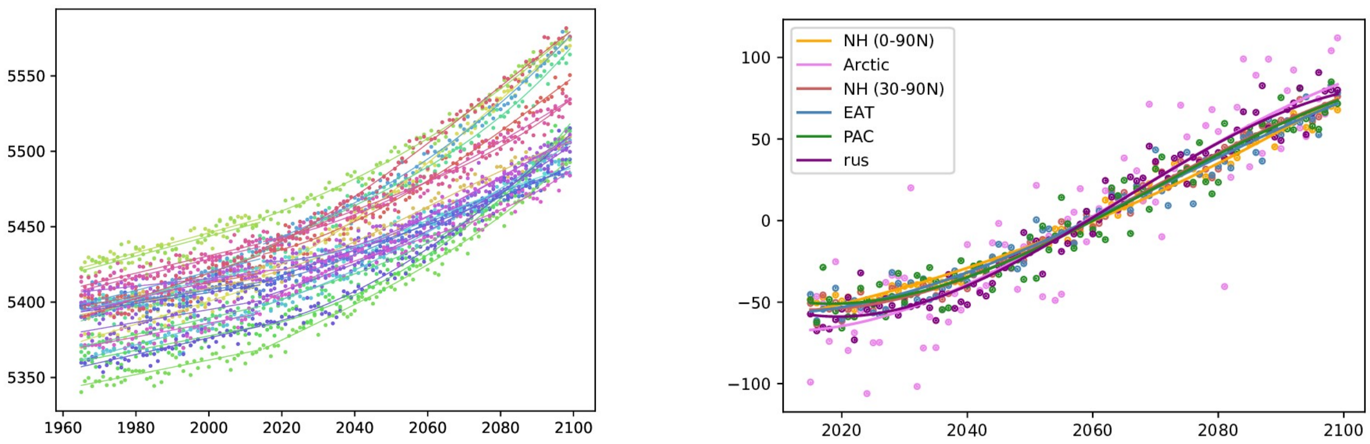

Figure 1. Average geopotential field in the Northern hemisphere (30-90N) for all model simulations in the historical+ssp585 scenario (scatter) and the linear/polynomial fit for the historical/scenario respectively (lines). Right panel: Detrending for the ssp585 scenario simulation with EC-Earth3, considering different areas for the average: NH (0-90N), Arctic (70-90N), NH (30-90N), EAT and PNA as in the paper, rus (30-90N, 40-140 E).

• 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)? 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 (now at lines 130-134): "*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 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 rephrased the paragraph to make this clearer (now at lines 142-144): "*The phase space spanned by these EOFs (hereafter "reference phase space") is then used for both the reanalysis and all GCMs simulations: all anomalies are projected onto this reference phase space, obtaining the 4 leading Principal Components (PCs) for the reanalysis dataset and 4 pseudo-PCs for each model simulation.*"

• 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 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 in cluster analysis. 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. We added a clarification at lines 170-172: "*In cluster analysis, a larger value of this ratio is generally desirable, indicating that the clusters are well separated from each other. For WRs, the distance from the observed variance ratio is an indicator of the overall model performance in simulating the regime dynamics (Fabiano et al., 2020).*"

• 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 produced an analogous figure with specific symbols for each model, and added it to the Supplementary material (Figure S1). 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 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 removed the claim on the two NAO regimes improving more than the other regimes. Lines 225-227 are changed accordingly: “*For the EAT sector (Figure 2, first row), the CMIP6 ensemble shows an improvement with respect to the CMIP5 counterpart for all regimes, although the intermodel spread is quite large and the ellipses significantly overlap (apart for NAO-).*”
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. We provided an indication of how many of the “matching” models improve from CMIP5 to CMIP6 in the pattern correlation, variance ratio and frequency bias in Table 2. The results confirm those of the main analysis, with most “matching” models improving from CMIP5 to CMIP6 for all metrics. Weaker improvements are found for the EAT regime frequency and for the PAC pattern correlations, as was also seen in the main analysis. Lines 244-245 now read: “*The results shown in Figures 2 and 3 are confirmed when looking at the performance of models developed by the same institution in the two phases, reported in Table 2: most models improve from*

CMIP5 to CMIP6 for all three 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?

- 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 added a brief comment on this in the paper at lines 231-233: “*PAC regimes appear to be more difficult to capture than EAT ones. This may reflect a larger natural variability in the observed regime structure, as suggested by the smaller variance ratio of the reanalysis for PAC with respect to EAT. Also, the PAC regime patterns might be influenced by the specific history of each model simulation in terms of amplitude and frequency of ENSO events.*“

• 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.
We added a brief comment on this at line 240: "*Given the strong link between WRs and blocking events (Madonna et al., 2017; Fabiano et al., 2020), the reduction of the WR frequency bias is in line with the smaller biases in the blocking frequency observed for CMIP6 models (Davini and D'Andrea, 2020).*"

• 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 (line 237-239) as follows: "*It is worth noting that – opposite to the EAT sector – models tend to produce larger variance ratios for the PAC regimes than it is observed, which might 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 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 S3, 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, the equivalent Figures with violin plots are available in the Supplementary Materials as Figure S5 and S8.

• 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.

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

• L257-263: The apparently non-linear response of the NAO- to the CO2 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. We added a brief comment on this at line 283: “*These might be related to the aerosol forcing, which has been hypothesized to have driven the observed AMV* *oscillations (Zhang et al., 2013; Qin et al., 2020). However, it is not clear whether a similar process might be at work for the future scenario period and further analysis on this topic will be carried out in a different study*.”

• 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 figure has been added to the supplementary material (Figure S10). 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. 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 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 added Figure 10 to the manuscript, which shows the change in the average number of regime events per 100 days. Also, we added lines 311-315: “*Figure 8 shows the number of regime events per 100 days. The changes in the regime frequencies might be seen as the combined effect of the changes in the regime persistence and the changes in the number of regime events. For the EAT sector, both have a comparable role in the frequency change of AR and SBL, while the increased persistence seems the main factor in the NAO+ change. For the PAC sector, the increase in PT frequency is driven by longer persistence, despite no significant change in the number of events, while the opposite is true for the change in the BR regime frequency*.”

• 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 added a comment on this in the discussion at line 347: “*A recent study of ENSO occurrence in CMIP6 projections (Fredriksen et al., 2020) shows a tendency for an increase of ENSO variability under global warming, and interestingly this change appears to be mostly related to an increase in positive El Niño events.*”

• 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 changed this to NH (30-90N) in the revised manuscript.

• Figure 9: Just for clarification, does the shading in this figure show the mean Z500 (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 NH trend. The difference of future and historical Z500 is dominated by the global NH 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 new figure has been added to the supplementary as Figure S12.

• 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 added the citation to the text.

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

- We added a couple of comments on this in the discussion at lines 329 and 342:
"*The strong increase in the NAO+ regime frequency is in line with the change of storm track activity in CMIP6 projections analyzed by Harvey et al. (2020), which shows an intensification of*

the 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. 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."
"*Also, a strong decrease of the storm track activity under SSP585 has been observed over the whole Northern American continent, and an increase in the central North Pacific (Harvey et al., 2020). This agrees well with the prediction of an increased PT regime, that blocks the entrance of perturbations in the continent.*"

• 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.

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

Anonymous Referee #2

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. Zappa & Shepherd 2017, Oudar et al. 2020), 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”.
Also, a comment on the fact that the links may be due, at least partly, to a reversed causal relationship or to an external forcing influencing both processes in a similar way, are included to the discussion in the revised manuscript at lines 392-395: “*This analysis aims at finding significant relationships between the set of potential drivers and the WR frequency trends. Of course, a significant correlation between the two quantities does not demonstrate the existence of a causal link - as they might be responding to a common external forcing - nor gives indication on the direction of such link. Nevertheless, this can provide an insight about the interconnection between mid-latitude climate variability and large scale changes in GCMs that deserve further investigation.*”
Also, we included a new comment in the discussion at lines 430-431, suggesting that it may also be that both the TNAW (tropical North-Atlantic warming) and the change in the Pacific WRs frequency share a common external driver: “*However, this significant regression may also be*

produced by a common external driver, influencing both the TNAW and the PNA regimes, like a change in the ENSO forcing (Fredriksen et al., 2020)."

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 North Atlantic (TNAW) and the North Atlantic suppolar gyre (SPGW). 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 CO_2 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 pseudo-PCs. 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 estimated the impact of the internal variability on our results by looking at two model ensembles for the historical and ssp585 simulations: MPI-ESM1-2-LR (10 members) and UKESM1-0-LL (4 members). The results for the EAT sector are shown in Figure 1 below: the running mean of the ensemble mean frequency, with ensemble spread (left panel), and the future change in WR frequency, with ensemble spread (right panel). The variability inside each model ensemble is smaller than that observed for the multi-model ensemble (Figure 4 of the manuscript), or at most of the same order, thus giving confidence in the results obtained using a single member for each model.

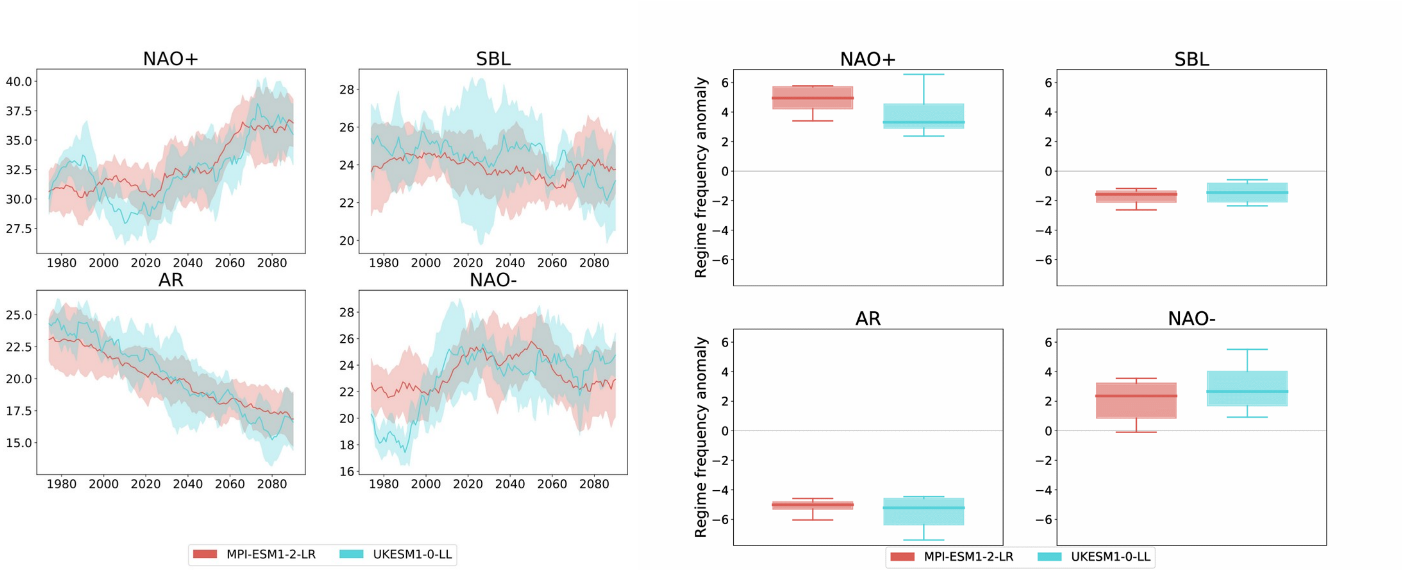

Figure 1. Estimation of the impact of the internal variability on the results, based on two multi-member ensembles. Left panel: running mean of the ensemble mean WR frequency for the two models, with indication of the ensemble spread (fill, standard deviation). Right panel: analog of Figure 4 of the manuscript showing the variability of the change in the WR frequency inside the ensemble.

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 at line 15: "*The wintertime mid-latitude climate in the Northern Hemisphere is primarily influenced by the low-frequency variability (at timescales longer than 5 days) related to the strength and position of the eddy-driven jet stream.*"

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

- Thank you for the suggestion, we moved the table to the main text (Table 1).

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 superseded 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 186 of the revised 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. The figure below (left panel) 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 exhibit 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 right panel in Figure 3). 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.

The two figures have been added to the supplementary material (Figure S11) and this point has been now clarified in the text at lines 123-126: "*We chose to calculate the trend on the Northern Hemisphere (30N-90N) - and not on the separate EAT and PAC domains - in order to retain possible decadal basin-wide fluctuations. Anyway, the difference in the future trends when considering the whole hemispheric or the sectorial averages is very small (see right panel in Figure S11).*"

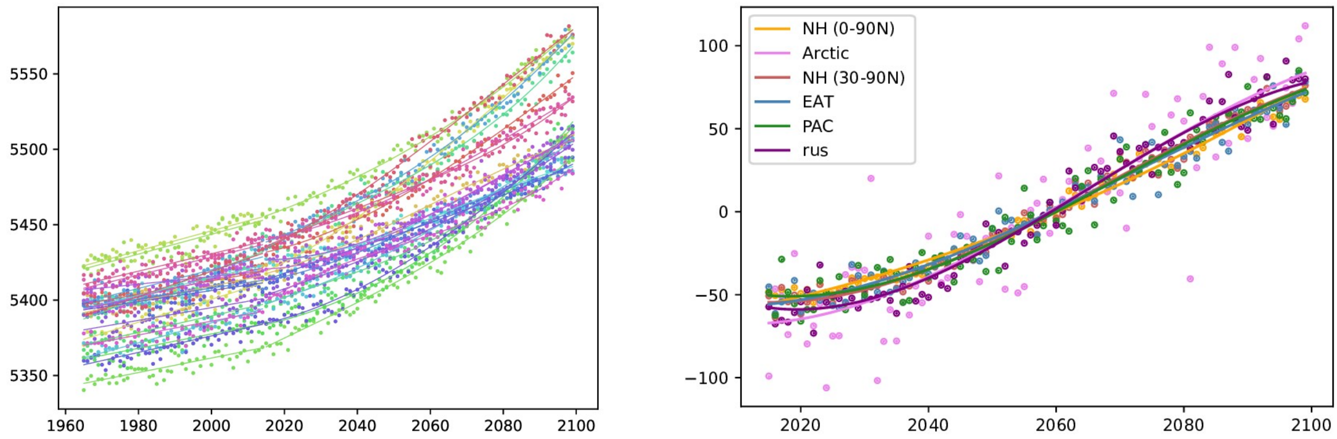

Figure 2. Left panel: average geopotential field in the Northern Hemisphere (30-90N) for all model simulations in the ssp585 scenario and corresponding historical simulation (scatter) and the linear/polynomial fit for the historical/scenario periods respectively (lines). Right panel: detrending for the ssp585 scenario simulation with EC-Earth3, considering different areas for the average.

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 at line 140, 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 at line 150: "[…] *we obtain a set of 4 cluster centroids, defined as the average of all PCs assigned to a certain cluster.*"

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.) which makes the methodology more arbitrary. 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 at line 186: "*The variability on a 50-yr window has been estimated as the standard deviation of the mean ($\sigma / \sqrt{n-1}$) of the seasonal frequency and persistence as 1.6% and 0.3 days respectively. However, the actual variability on these scales might be larger than this due to decadal basin-wide fluctuations.*"

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 (lines 237-239): "*It is worth noting that – opposite to the EAT sector – models tend to produce larger variance ratios for the PAC regimes than it is observed, which might 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.
We added a comment on this at lines 283-285: "*These might be related to the aerosol forcing, which has been hypothesized to have driven the observed AMV oscillations (Zhang et al., 2013; Qin et al., 2020). However, it is not clear whether a similar process might be at work for the future scenario period and further analysis on this topic will be carried out in a different study.*"

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 has been added to the Supplementary material (Figure S11, left panel). The residual zonal and local trends of zg500 for ssp585 are shown in Figures 9 and 10. 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, this is indicated at line 355.

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 predictors, divided by the global mean surface temperature trend. Trends are calculated on the 2015-2100 period for ssp585 and on 2006-2100 for rcp85. We added this information to the main text at line 388.

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 computed the correlations between the frequency trends for ssp585 in the two sectors. No significant correlation at the 95% level was found. The largest correlations are: AR and BR (+0.3), NAO- and PNA- (-0.3), SBL and PT (-0.3), SBL and PNA- (+0.35).

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 predictors 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 indices 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. However we also added here the 19 models from RCP8.5, to enlarge the sample to 38 models.

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, which is now included as Figure S12. The significance of the correlations at 99% and 95% level is indicated by the big and small white circles in Figure 11. The issue of the effective sample size might effectively reduce the significance of these correlations, we added a brief comment on this in the revised manuscript at line 404: "*However, the effective size of the sample may be smaller than the 38 models considered here (19 for both SSP5-8.5 and RCP8.5), since some models are closely related to each other and this might lower the significance of the regressions.*"

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.

A regime view of future atmospheric circulation changes in Northern mid-latitudes

Federico Fabiano[1], Virna L. Meccia[1], Paolo Davini[2], Paolo Ghinassi[1], and Susanna Corti[1]

[1]Institute of Atmospheric Sciences and Climate (ISAC-CNR), Bologna, Italy
[2]Institute of Atmospheric Sciences and Climate (ISAC-CNR), Turin, Italy

Correspondence: F. Fabiano (f.fabiano@isac.cnr.it)

Abstract. Future wintertime atmospheric circulation changes in the Euro-Atlantic (EAT) and Pacific-North American (~~PNA~~PAC) sectors are studied from a ~~Weather Regimes~~ weather regimes perspective. The CMIP5 and CMIP6 historical simulations performance in reproducing the observed regimes is first evaluated, showing a general improvement of CMIP6 models, more evident for EAT. The circulation changes projected by CMIP5 and CMIP6 scenario simulations are ~~analyzed~~ analysed in terms of the change in the frequency and persistence of the regimes. In the EAT sector, significant positive trends are found for the frequency and persistence of NAO+ for ~~SSP245, SSP370 and SSP585~~ SSP2-4.5, SSP3-7.0 and SSP5-8.5 scenarios, with a concomitant decrease in the frequency of the Scandinavian Blocking and Atlantic Ridge regimes. For ~~PNA~~PAC, the Pacific ~~Through~~ Trough regime shows a significant increase, while the Bering Ridge is predicted to decrease in all scenarios ~~analyzed~~analysed. The spread among the model responses is linked to different levels of warming in the ~~Polar Stratosphere~~polar stratosphere, the tropical upper troposphere, the North Atlantic and the Arctic.

1 Introduction

A major challenge for the climate community is ~~understanding~~ to understand how a warmer climate will affect the large-scale atmospheric circulation in the mid-latitudes. Indeed, there is growing interest on this topic both from the scientific community and from society, as future changes in circulation are also inextricably related ~~with~~ to regional impacts and the occurrence of extreme weather conditions (e.g. Brunner et al., 2018; Schaller et al., 2018; Screen and Simmonds, 2014; Sousa et al., 2018). The wintertime mid-latitude climate in the Northern Hemisphere is primarily influenced by the low-frequency variability (at timescales longer than 5 days) related to the strength and position of the eddy-driven jet stream (Woollings et al., 2010; Barnes and Polvani, 2013). This is particularly true for the North-Atlantic and North-Pacific sectors, where the latitudinal shifts of the jet describe a significant fraction of the low-frequency variability (Athanasiadis et al., 2010) and determine specific impacts locally (Ma et al., 2020) and over downstream regions (i.e. Europe and North America) (Screen and Simmonds, 2014; Zappa et al., 2015a, b; Loikith and Broccoli, 2014). In recent years, several studies on the future evolution of the mid-latitude atmospheric circulation focused on the changes in the mean state of the jet streams, mainly in terms of latitudinal shifts and changes in the jet speed (Barnes and Polvani, 2013). Further attention on the topic has grown in the last decade following the debate about the influence that the Arctic ~~Amplification~~ amplification - i.e. the faster warming of the surface at high latitudes

- may have on the jet structure (Barnes and Screen, 2015; Hoskins and Woollings, 2015). The emerging picture is that the fate of the eddy-driven jet streams in a warmer climate is mainly controlled by the meridional temperature gradient in mid-latitudes, which in turn depends on three independent processes, all linked to the differential heating of different regions of the atmosphere:

- the faster warming of the tropical upper troposphere - known as ~~Upper Tropospheric Warming~~ upper tropospheric warming (UTW) -, mainly driven by increased convection and upper-level latent-heat release (Peings et al., 2017). The effect of UTW is to increase the meridional temperature gradient in the upper troposphere and promote a poleward shift and intensification of the jet ~~(Hassanzadeh et al., 2014)~~(Barnes and Screen, 2015).

- The Arctic ~~Amplification~~ amplification (AA), which is due primarily to sea-ice retreat and increased heat flux from the ocean in autumn and winter, along with several other positive feedbacks in the Arctic region ~~(Screen and Simmonds, 2010)~~ (Screen and Simmonds, 2010; Pithan and Mauritsen, 2014). The effect of AA is to decrease the low-level meridional temperature gradient and to promote a slow down and equatorward shift of the jet (Peings et al., 2017; Hassanzadeh et al., 2014; Cohen et al., 2019; Overland et al., 2016).

- The change of the ~~Polar Stratospheric Temperature~~ polar stratospheric temperature (PST) and consequent feedback on the ~~Polar Vortex Strength~~ polar vortex strength (PVS). The fate of the polar stratosphere is still unclear, but there is some indication that it may be of primary importance for the North-Atlantic jet stream (Manzini et al., 2014; Zappa and Shepherd, 2017; Peings et al., 2017).

Due to the large internal variability of the system on interannual to decadal time-scales, detecting circulation changes in the observations has proven to be a challenging task and multiple circulation indices do not show significant trends during the observational period (Blackport and Screen, 2020; Barnes and Screen, 2015). Nevertheless, some robust indications of future changes come from General Circulation Models (GCMs) simulations under greenhouse gases (GHG) forcing scenarios. Analyses on the Coupled Model Intercomparison Project - Phase 5 (CMIP5 Taylor et al., 2012) and Phase 6 (CMIP6 Eyring et al., 2016) ensembles have shown a general agreement for a moderate (about 1 deg) northward shift of the annual-mean zonal-mean eddy-driven jet by 2100 (Barnes and Polvani, 2013; Shaw et al., 2016). However, the picture appears more complex than that, and the jet response strongly depends on the region (Barnes and Polvani, 2013; Peings et al., 2017) and season considered (Barnes and Polvani, 2015; Shaw et al., 2016). Whilst the northward shift of the jet is evident in the North-Pacific (Oudar et al., 2020), the trend over the North-Atlantic shows rather a squeezing of the time-mean jet, with intensification and eastward elongation of the westerlies over Europe (Oudar et al., 2020; Peings et al., 2018). The eastward elongation of the North-Atlantic jet is also consistent with results obtained from analysis of the stationary waves response to climate change, which show an eastward shift in phase, produced by a decrease in the stationary zonal wavenumber (Wills et al., 2019; Simpson et al., 2014). Also from the dynamical point of view, the response in the North-Atlantic appears to be more complex, with concurrent and opposite influences of UTW and AA (Peings et al., 2017). This picture is further complicated by the emerging role of the polar stratosphere, which is strongly coupled with the North-Atlantic jet stream position (e.g. Baldwin and Dunkerton, 2001), and might contribute to the dynamical response to the UTW (Peings et al., 2017; Manzini et al., 2014; Beerli and Grams, 2019), therefore explaining the large inter-model spread (Peings et al., 2017; Zappa and Shepherd, 2017; Oudar et al., 2020).

Besides changes in the mean state, a large interest has been given to changes in the day-to-day variability of the jets, also motivated by the fact that climate extremes are commonly related with persistent circulation anomalies. Barnes and Polvani (2013) showed that, in the North-Atlantic, there is a decreasing trend in the first mode of variability of the jet – related to the latitudinal shifts – and an increase in the second mode – related to variations in jet speed – under RCP8.5 scenario. This picture has been confirmed by the analysis by Peings et al. (2018), which claims that there would be less room for latitudinal shifts of the jet due to the squeezing produced by UTW and AA. Analyses based on various indices of "waviness", i.e. of the departure from a purely zonal jet structure, predict a more zonal flow under a warmer climate (Blackport and Screen, 2020; Peings et al., 2017), with the possible exception of the North-American region (Di Capua and Coumou, 2016; Peings et al., 2017; Vavrus et al., 2017). This is also confirmed by analysis of the atmospheric blocking frequencies in CMIP5 and CMIP6 models, with a general decrease in winter blocking over the Northern Hemisphere and a tendency to an eastward shift of the blocking maxima, with the only small (non-significant) increase over western Canada (Davini and D'Andrea, 2020; Woollings et al., 2018).

In this work, we propose an alternative view of future changes in the wintertime circulation at Northern mid-latitudes, based on the analysis of the daily geopotential height at 500 hPa. With respect to the climatological reference state, midlatitude disturbances appear as transient geopotential height anomalies that can persist beyond the typical synoptic scale, up to three or four weeks. In some regions, the flow tends to organize in some preferred configurations, although the number of such configurations to be considered is still a matter of debate (Hannachi et al., 2017). These preferred large-scale circulation patterns are commonly referred to as ~~Weather Regimes~~ weather regimes (WRs) and have been studied mostly for the Euro-Atlantic ~~(EAT Dawson et al., 2012; Strommen et al., 2019)~~ (EAT: Michelangeli et al., 1995; Dawson et al., 2012; Strommen et al., 2019) and Pacific-North American ~~(PNA Straus et al., 2007; Weisheimer et al., 2014)~~ (PAC: Straus et al., 2007; Weisheimer et al., 2014) sectors. Each WR has a different impact on the climate of the ~~downstream region~~ region considered, driving specific precipitation and temperature anomalies. Many works in literature studied how WRs are reproduced by GCMs, but mostly focused on the model performance in control or historical simulations (Dawson et al., 2012; Cattiaux et al., 2013b; Dawson and Palmer, 2015; Weishei . Changes of WRs in CMIP5 projections were analysed by Cattiaux et al. (2013a) and by Ullmann et al. (2014) for the EAT sector.

From an heuristic perspective, the WRs can be seen as the attractors of a nonlinear dynamical system, whose main characteristics may be described in terms of their position in phase space and their frequency of occurrence. In simple dynamical systems, under a small external forcing the main structure of the attractors in the phase space is only marginally affected by the forcing (at least at the first order), while it is the frequency of occurrence of the regimes that changes in response to the forcing, with some regimes becoming more populated (Palmer, 1999). By analogy, a similar response to forcing has been hypothesized for the WRs in complex GCMs (Palmer, 1999; Corti et al., 1999). Here we use this framework to evaluate the change in the frequency of occurrence of the WRs in the future scenarios, as simulated by the climate models participating in CMIP5 and CMIP6.

The paper is structured as follows: Section 2 describes the Data and Methods used for the analysis; Section 3 shows the results regarding the observed WRs, the model performance and the future projections for the EAT and ~~PNA~~ PAC sectors; in the Discussion (Section 4) the results are commented with respect to changes in the climate mean state and ~~analyzing~~ analysing the connection between the multi-model spread and multiple drivers of the circulation changes; the conclusions are summarized in Section 5.

2 Data and methods

2.1 Data

An ensemble of GCMs simulations being part of the Coupled Model Intercomparison Project - Phase 5 (CMIP5 Taylor et al., 2012) and Phase 6 (CMIP6 Eyring et al., 2016) are here ~~analyzed~~analysed. For CMIP6, we considered both the historical and four future scenarios simulations with different levels of anthropogenic carbon dioxide emissions along the 21st century: SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5 (O'Neill et al., 2016). The SSP1-2.6 scenario corresponds to significantly reduced fossil fuel burning by mid-century and a global warming contained at about 2° C, while the SSP5-8.5 is the business-as-usual scenario. SSP2-4.5 and SSP3-7.0 are intermediate scenarios (Meinshausen et al., 2019). For CMIP5, we consider the historical and the most extreme future scenario RCP8.5 (Moss et al., 2010), which features a smaller CO_2 concentration by 2100 than SSP5-8.5 but larger than SSP3-7.0 (Meinshausen et al., 2019). We included 33 CMIP6 and 27 CMIP5 models in the analysis of the model performance for the historical simulations (Section 3.2). The results regarding the future scenarios (Section 3.3) are restricted to the models that were available both in the historical simulation and in all future scenarios considered, resulting in 19 models for both CMIP5 and CMIP6. The models and ensemble members used in the analysis are listed in ~~the Supplementary material, Table S1~~Table 1.

The reference period for historical simulations spans from 1964 to 2014 for CMIP6 (1964 to 2005 for CMIP5). However, the common period 1964-2005 is considered when comparing the performance of the two ensembles (Section 3.2). The reanalysis data from a combination of the European Centre for Medium-Range Weather Forecast (ECMWF) ERA-40 (1964-1978 Uppala et al., 2005) and ERA-Interim (1979-2014 Dee et al., 2011) is used as a reference. Selecting a different reanalysis product as NCEP does not affect the results, as discussed in Fabiano et al. (2020). For the computation of the ~~Weather Regimes~~ weather regimes we consider the wintertime (NDJFM) daily mean geopotential height at 500hPa ("data" in the following). For practical reasons, data are first interpolated to a 2.5° x 2.5° grid using a bilinear interpolation. As discussed in Fabiano et al. (2020), since we are interested in the large scale patterns, this does not impact on results. Since in Section 4.2 we evaluate the role of some drivers in determining the inter-model spread of the response in SSP5-8.5 and RCP8.5 scenarios, monthly averages of the atmospheric temperature (ta) at different vertical levels and wind (ua) in the stratosphere are also used.

Table 1. Models and ensemble members used in the analysis. Hist stands for historical simulation and ssps means that SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5 were all analysed for that model/member (for CMIP6).

CMIP5			CMIP6		
Model	Member	Experiment	Model	Member	Experiment
ACCESS1-0	r1i1p1	hist, RCP8.5	ACCESS-CM2	r1i1p1f1	hist, ssps
ACCESS1-3	r1i1p1	hist, RCP8.5	AWI-ESM-1-1-LR	r1i1p1f1	hist
BNU-ESM	r1i1p1	hist, RCP8.5	BCC-CSM2-MR	r1i1p1f1	hist, ssps
CMCC-CESM	r1i1p1	hist, RCP8.5	BCC-ESM1	r1i1p1f1	hist
CMCC-CM	r1i1p1	hist, RCP8.5	CanESM5	r1i1p1f1	hist, ssps
CMCC-CMS	r1i1p1	hist, RCP8.5	CESM2-FV2	r1i1p1f1	hist
CNRM-CM5	r1i1p1	hist, RCP8.5	CESM2	r1i1p1f1	hist
CanESM2	r1i1p1	hist, RCP8.5	CESM2-WACCM-FV2	r1i1p1f1	hist
FGOALS-g2	r1i1p1	hist, RCP8.5	CESM2-WACCM	r1i1p1f1	hist, ssps
GFDL-CM3	r1i1p1	hist, RCP8.5	CNRM-CM6-1-HR	r1i1p1f2	hist, ssps
GFDL-ESM2G	r1i1p1	hist	CNRM-CM6-1	r1i1p1f2	hist, ssps
HadGEM2-CC	r1i1p1	hist, RCP8.5	CNRM-ESM2-1	r1i1p1f2	hist, ssps
HadGEM2-ES	r1i1p1	hist	EC-Earth3	r1i1p1f1	hist, ssps
IPSL-CM5A-LR	r1i1p1	hist	FGOALS-f3-L	r1i1p1f1	hist
IPSL-CM5A-MR	r1i1p1	hist	FGOALS-g3	r1i1p1f1	hist, ssps
IPSL-CM5B-LR	r1i1p1	hist	GFDL-CM4	r1i1p1f1	hist
MIROC-ESM-CHEM	r1i1p1	hist, RCP8.5	GISS-E2-1-G	r1i1p1f1	hist
MIROC-ESM	r1i1p1	hist, RCP8.5	GISS-E2-1-G	r1i1p1f2	hist
MIROC5	r1i1p1	hist, RCP8.5	HadGEM3-GC31-LL	r1i1p1f3	hist
MPI-ESM-LR	r1i1p1	hist	HadGEM3-GC31-MM	r1i1p1f3	hist
MPI-ESM-MR	r1i1p1	hist, RCP8.5	INM-CM4-8	r1i1p1f1	hist, ssps

2.2 Methods

2.2.1 Trend and seasonal cycle removal

The geopotential height field shows a clear increasing trend both in the historical and scenario simulations, due to global warming (see Figure S11). Before computing the ~~Weather Regimes~~weather regimes, data are detrended by applying two different methodologies. Data from historical simulations and reanalysis are detrended by removing the linear trend of the area-weighted season-averaged Northern Hemisphere (30-90N) geopotential height time series. We chose to calculate the trend on the Northern Hemisphere (30N-90N) - and not on the separate EAT and PAC domains - in order to retain possible decadal basin-wide fluctuations. Anyway, the difference in the future trends when considering the whole hemispheric or the sectorial averages is very small (see right panel in Figure S11). For future scenarios, the detrending is implemented by fitting a third-order polynomial to the above mentioned Northern Hemisphere time series and removing this from the geopotential height field: this is done to suitably fit the acceleration in the global increase in geopotential height seen in the second half of the century. Once the trends are removed, the mean seasonal cycle is subtracted from the data to obtain detrended daily geopotential height anomalies ("anomalies" in the following). 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. It is worth noting that the above-defined average seasonal cycle computed in the historical simulations might differ 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, the mean seasonal cycle is computed in the reference period of the historical simulation (1964-2014 for CMIP6, 1964-2005 for CMIP5).

2.2.2 Weather ~~Regimes~~ regimes computation

The ~~Weather Regimes~~ weather regimes are computed using the WRtool Python package (Fabiano et al., 2020). We focus here on latitudes between 30° and 90° N and consider separately two longitudinal sectors: the Euro-Atlantic (EAT, from 80°W to 40°E) and the Pacific-North American (~~PNA~~PAC, from 140°E to 80°W). The procedure is identical for the two sectors. To reduce dimensionality, an Empirical Orthogonal Function (EOF) decomposition is applied to the observed anomalies, retaining the 4 leading EOFs, which explain 53% and 48% of the total variance for the EAT and ~~PNA~~ PAC sectors respectively. ~~Both anomalies from reanalysis and GCMs simulations are then projected into the~~ Sensitivity tests performed in Fabiano et al. (2020) for the EAT sector show that the changes in the regime patterns when considering for example 10 EOFs instead of 4 are negligible. The phase space spanned by these EOFs (hereafter "reference phase space") ~~. In this way,~~ is then used for both the reanalysis and all GCMs simulations: all anomalies are projected onto this reference phase space, obtaining the 4 leading Principal Components (PCs) ~~from~~ for the reanalysis dataset and 4 pseudo-PCs ~~from~~ for each model simulation~~are obtained. The Weather Regimes~~ . The weather regimes for the reanalysis are computed by applying a K-means clustering algorithm to the PCs. For the EAT sector, we set the number of regimes to 4, as is widely documented in literature (Michelangeli et al., 1995; Cassou, 2008; Dawson et al., 2012; Madonna et al., 2017; Strommen et al., 2019; Fabiano et al., 2020). For the ~~PNA~~ PAC sector, we choose 4 clusters as in Straus et al. (2007) and Weisheimer et al. (2014), although a different number of clusters could be a viable alternative as argued in Straus et al. (2007) and favoured by other studies (e.g. Kimoto and Ghil, 1993; Michelangeli et al., 1995; Robertson and Ghil, 1999). Each day is assigned to one of the regimes and we obtain a set of 4 cluster centroids, defined as the average of all PCs assigned to a certain cluster. The cluster centroids obtained for the reanalysis are referred to as "reference centroids" in the following. The regime pattern is defined as the composite of all anomalies assigned to a certain regime.

For the models, we follow two different approaches to assign each day to a specific regime and, accordingly, two regime types are defined:

- *Computed regimes*: the K-means clustering is performed on the pseudo-PCs and 4 simulated cluster centroids are obtained, as in Fabiano et al. (2020). These *computed regimes* are calculated for the historical simulations only, in order to compare observed and simulated regimes structure and to assess possible model deficiencies.

- *Projected regimes*: the K-means algorithm is not applied, but each anomaly in the reference phase space is attributed to the closest reanalysis reference centroid. In this way, the regimes are consistently defined for all simulations and the variability in the clustering itself as a possible source of noise is ruled out. The *projected regimes* are used to compare the regime frequencies and persistence across different simulations/scenarios within a common reference framework.

2.2.3 Metrics

Here below a set of metrics used in Section 3 is defined:

- *Taylor diagram.* A Taylor diagram (Taylor, 2001) is used as a synthetic metric to evaluate how the simulated regime patterns resemble the observed ones. The Taylor diagram consists of a polar plot showing the spatial correlation between the simulated and observed patterns (angular axis) and their standard deviation (radial axis, in units of the observed standard deviation). Due to the geometrical construction, the linear distance between the simulation and the observation is the centered-pattern RMS (with bias subtracted).

- *Variance ratio.* The variance ratio is defined as the ratio of the average inter-cluster squared distance to the mean intra-cluster variance. ~~It is a measure of the geometrical regime structure and it~~ In cluster analysis, a larger value of this ratio is generally desirable, indicating that the clusters are well separated from each other. For WRs, the distance from the observed variance ratio is an indicator of the overall model performance in simulating the regime dynamics (Fabiano et al., 2020).

- *Regime frequency.* The regime frequency over a certain period is defined as the fraction of days assigned to a certain regime in that period. Accordingly, the seasonal regime frequency is a time series indicating the fraction of days assigned to a certain regime in each season. In order to estimate each model performance, the "absolute frequency bias" is defined as the absolute difference between the simulated and observed regime frequency, averaged over all regimes.

– *Regime persistence.* The regime persistence is the average duration in days of a given regime event. A regime event is a set of consecutive days assigned to the same regime. We relaxed this definition to allow for single day departures from the regime state, thus a regime event is ended only when two consecutive days are assigned to different regimes.

The Taylor diagram, the variance ratio and the bias in regime frequency – calculated for the computed regimes of the historical simulations – are used to evaluate the ability of climate models in reproducing ~~Weather Regimes~~ weather regimes (Sections 3.2). The change of the projected regimes' frequencies and persistence in future scenarios will be ~~analyzed~~ analysed in Section 3.3.

The observed interannual variability of the regime frequencies and persistence has been estimated to be about 11% and 2.5 days respectively, averaged over all regimes. Therefore, in order to assess significant long-term changes we average these quantities over the following periods: 1964-2014 for the CMIP6 historical runs (1964-2005 for CMIP5) and 2050-2100 for the scenarios. The variability on a 50-yr window ~~is about~~ has been estimated as the standard deviation of the mean $(\sigma/\sqrt{n-1})$ of the seasonal frequency and persistence as 1.6% and 0.3 days ~~for the frequency and persistence respectively, assuming that consecutive seasons are statistically independent~~respectively. However, the actual variability on these scales might be larger than this due to decadal basin-wide fluctuations. For the frequencies, in addition to the differences between the scenario and historical simulations in specific time-windows, the trends in the 2015-2100 period of the scenarios are computed. In order to do this, a 10-year running mean to the time series is applied.

3 Results

3.1 Observed regimes

The regime centroids obtained from the reanalysis (NDJFM, 1964-2014) for the Euro-Atlantic sector and the Pacific-North America are shown in Figure 1. For the EAT sector, the four regimes are the two phases of the North Atlantic Oscillation (NAO+, NAO-), the Scandinavian Blocking (SBL) and the Atlantic Ridge (AR). The patterns are consistent with those obtained considering different periods and using a different definition for boreal winter (i.e. DJF or DJFM) (Dawson et al., 2012; Fabiano et al., 2020; Cassou, 2008). A close correspondence exists among these regimes, the structure of the North Atlantic jet stream and climatic conditions over Europe. The positive (negative) NAO is related to a central (southern) jet position, whereas a northward displacement of the jet is linked to the Atlantic Ridge regime (Madonna et al., 2017; Fabiano et al., 2020). The SBL regime is related to a high pressure over Scandinavia and corresponds to a tilted jet structure from SW to NE.

The four regimes in the Pacific sector are: the Pacific ~~Through~~ Trough (PT) (Straus et al., 2007) - the Rockies Ridge in Casola and Wallace (2007) -, the positive and negative phases of the Pacific-North American patterns (PNA+, PNA-; e.g. Wallace and Gutzler, 1981; Barnston and Livezey, 1987), and the Bering Ridge (BR) - also known as Alaskan Ridge - characterised by a blocked flow (Renwick and Wallace, 1996; Smyth et al., 1999; Straus et al., 2007; Casola and Wallace, 2007). The observed regime patterns over the Pacific are consistent with those found by Casola and Wallace (2007) and Weisheimer et al. (2014), although these authors considered different periods and data sets. All four WRs over the ~~PNA~~ PAC sector can

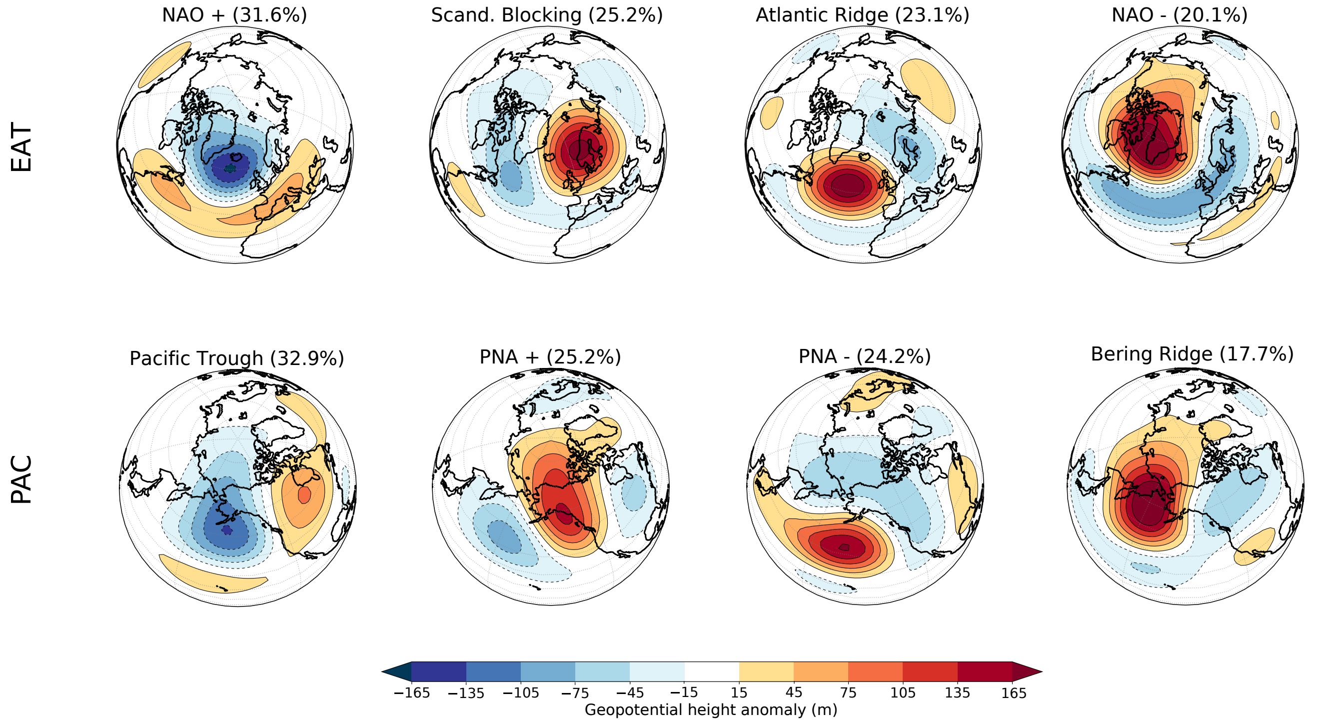

Figure 1. Regime patterns for the Euro-Atlantic (top row) and Pacific-North American regimes (bottom row), obtained from the reanalysis (1964-2014, NDJFM). The observed regime frequencies are indicated in the subplot titles. Note that the projection is centered in the Atlantic (Pacific) for EAT (~~PNA~~PAC) regimes.

be seen as different phases of a Rossby wave train extending from the Pacific towards the North American continent. The variability of such quasi-stationary patterns is modulated by both the interaction of the midlatitude jet with the orography and the forcing by the convection over the equatorial and tropical Pacific, which acts as a Rossby wave source (Trenberth 1998). The PNA+ pattern is in fact in phase with the barotropic response obtained from the interaction of the westerly jet with the orographic forcing provided by the Rocky mountains, and therefore it is associated with an enhanced ridge-through pattern over North America. The PT regime, which is strongly correlated with positive ENSO (Straus et al., 2007; Casola and Wallace, 2007; Weisheimer et al., 2014), exhibits an eastward shift compared to the PNA+. This eastward shift is related to the upper-tropospheric divergence caused by the enhanced convection over the Pacific during the positive ENSO events, which acts as an additional thermal Rossby wave source (Straus and Shukla, 2002). The PNA- and BR patterns appear as out of phase with the PNA+ and PT respectively and have been found to be correlated with La Niña events (Straus et al., 2007; Weisheimer et al., 2014).

3.2 Models' performance

In this subsection the model performance in reproducing the observed ~~Weather Regimes~~ weather regimes is assessed. The computed regimes of the historical simulations (for CMIP5 and CMIP6 models) are considered, and the models' performance is evaluated in terms of the regime centroids, regime frequency bias and variance ratio. The results are shown in Figures 2 and 3. Also, a complementary indication of the relative performance of CMIP5 and CMIP6 models is given in Table 2, which shows the number of models developed by the same institution that improve from CMIP5 to CMIP6 for the three metrics.

Figure 2 displays a set of Taylor diagrams in which the simulated regimes are compared with the observed ones: each simulation is shown by a dot and the overall performance of each ensemble is indicated by the shaded ellipses centered at the ensemble average and with semi-axes equal to the ensemble standard error. For the EAT sector (Figure 2, first row), the CMIP6 ensemble shows an improvement with respect to the CMIP5 counterpart for all regimes~~. The improvement is larger for the two NAO regimes, while it is seen mostly in terms of standard deviation for SBL and AR. For the PNA~~ , although the intermodel spread is quite large and the ellipses significantly overlap (apart for NAO-).

For the PAC regimes (Figure 2, second row), the difference in the overall performance of the two ensembles is less evident. General improvements are seen in CMIP6 for the negative PNA regime, and, in terms of standard ~~deviation~~error, for the PT and BR regimes. The performance in simulating the positive PNA regime is comparable in the two ensembles, with a slight worsening in CMIP6 in terms of standard ~~deviation~~error. PAC regimes appear to be more difficult to capture than EAT ones. This may reflect a larger natural variability in the observed regime structure, as suggested by the smaller variance ratio of the reanalysis for PAC with respect to EAT. Also, the PAC regime patterns might be influenced by the specific history of each model simulation in terms of amplitude and frequency of ENSO events.

The model performance in reproducing the observed variance ratio and regime frequency is shown in Figure 3. ~~CMIP6 models perform better than CMIP5 ones, with the ellipse getting closer to the observed value (black star) for both EAT and PNA regimes. The improvement is clearer in terms of variance ratio, but also the spread in the regime frequencies is reduced.~~ For CMIP6 (CMIP5), the average absolute frequency bias is about 2.2% (2.5%) over EAT and 2.3% (2.8%) over ~~PNA~~PAC, while the variance ratio is around 0.74 (0.72) for EAT and 0.76 (0.79) for ~~PNA. It 's~~ PAC. For the variance ratio (left panel), CMIP6 models perform better than CMIP5 ones, with the box getting closer to the observed value (black star) for both EAT and PAC regimes. It is worth noting that – opposite to the EAT sector – models tend to produce larger variance ratios for the ~~PNA~~ PAC regimes than it is observed~~. This may be linked to the fact that (some of) the PNA regime frequencies are modulated by the tropical Pacific forcing, that is very well simulated by state of the art climate models (Molteni et al., 2020)~~, which might be due to an excess in the tropically-induced modulation of the PAC regimes in models. An improvement in CMIP6 is also seen for the frequency bias (right panel), more evident in the PAC sector, but also detectable in the reduction of the spread of the EAT sector. Given the strong link between WRs and blocking events (Madonna et al., 2017; Fabiano et al., 2020), the reduction of the WR frequency bias is in line with the smaller biases in the blocking frequency observed for CMIP6 models (Davini and D'Andrea, 2020).

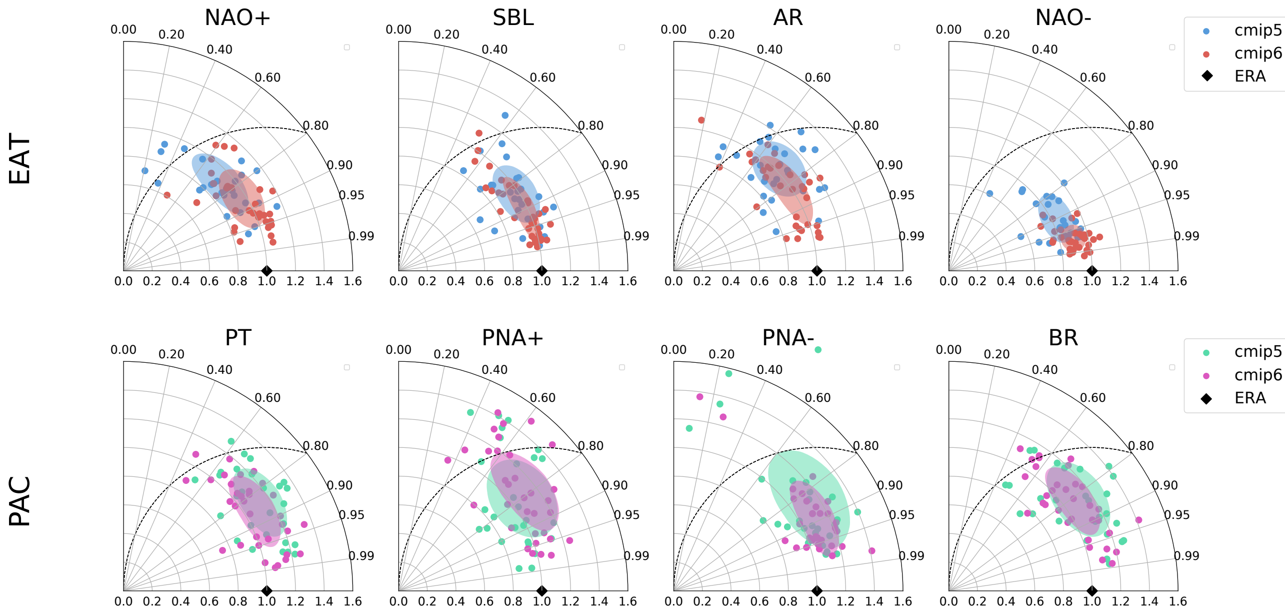

Figure 2. Taylor diagrams showing the agreement between simulated and observed regime patterns for CMIP5 (blue and green) and CMIP6 (red and pink) models. The shaded ellipses are used to indicate the overall ensemble performance: they are centered at the ensemble mean and have semi-axes equal to the ensemble standard error The simulated patterns are those obtained from the computed regimes of the historical simulations in the common period 1964-2005. The observed patterns are those computed from the reanalysis.

Table 2. Number of models developed by the same institution that improve from CMIP5 to CMIP6 for three metrics: average pattern correlation, average frequency bias and variance ratio. There is a total of 11 institution for which at least one model is available for both CMIP phases. If more model versions are available for a single institution, the average metric among all models is used.

Metric	EAT	PAC
Pattern	9/11	6/11
Freq. bias	6/11	9/11
Var. ratio	10/11	9/11

The results shown in Figures 2 and 3 are confirmed when looking at the performance of models developed by the same institution in the two phases, reported in Table 2: most models improve from CMIP5 to CMIP6 for all three metrics.

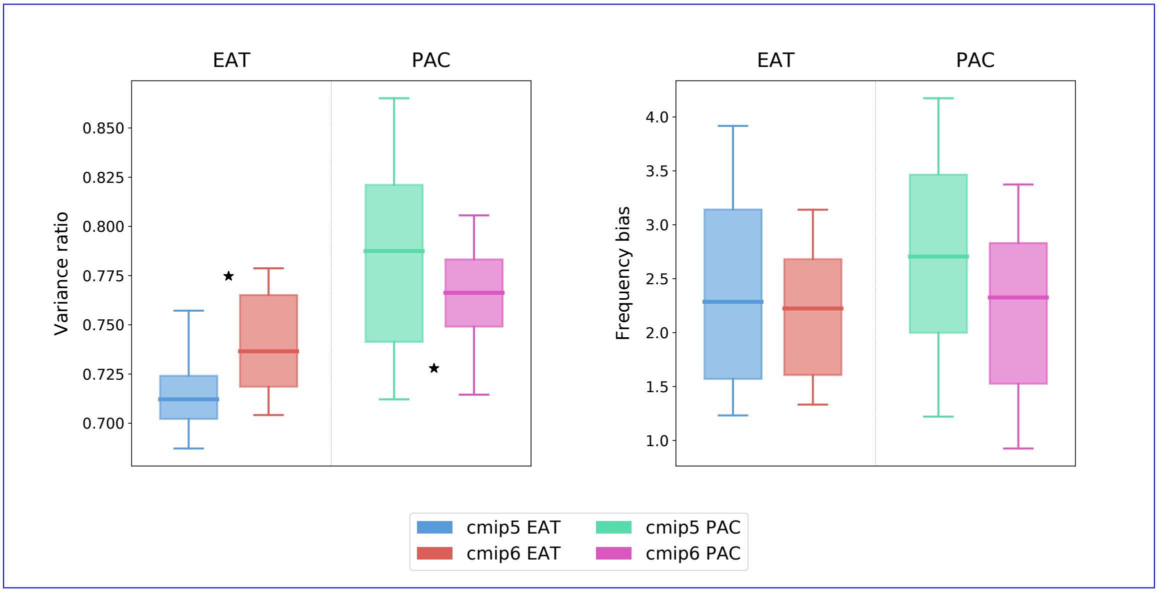

Figure 3. Variance ratio (~~x-axis~~left panel) and absolute frequency bias (~~y-axis~~right panel) for EAT ~~regimes (left panel)~~ and ~~PNA~~ PAC regimes~~(right panel)~~. CMIP5 models are indicated by blue (green for ~~PNA~~PAC) dots and CMIP6 by red (pink for ~~PNA~~PAC) ~~dots~~boxes. The ~~ellipses show the average ensemble performance: they are centered at~~ observed values of the ~~ensemble mean and have semi-axes equal to~~ variance ratio in the ~~ensemble standard error. The observed value is~~ two sectors are indicated by the black ~~star at the lower right corner~~stars.

3.3 Future scenarios

3.3.1 Regime frequency

We ~~analyze~~ analyse here the changes in the regime frequencies in the future scenarios. As discussed in Section 2.2.2, the frequency of the projected regimes is considered here, i.e. those computed by attributing models data to the closest reference centroid. By doing so we avoid that changes in frequency are (even partially) produced by a change in the pattern. Also, since the model biases in reproducing the observed regime frequencies and patterns are significantly smaller for projected regimes than for computed regimes (see Supplementary materials, Figures ~~S1 and~~ S2 and S3), this choice might lead to a somehow higher confidence in terms of future projections.

The difference between the regime frequencies for the future and the historical reference periods is shown in the top panels of Figure 4. Each panel shows whiskers plots representing the multi-model distribution of the regime frequencies for the reference periods (1964-2014 for CMIP6 historical, 1964-2004 for CMIP5 historical, and 2050-2100 for the scenarios). The CMIP5 historical frequency is shown in Figure ~~S3~~ S4 and the linear trends of the regime frequencies for the 2015-2100 period of the scenarios are shown in Figure ~~S4~~ S5 in the Supplementary Materials. Figure 5 shows the ensemble mean of the WRs seasonal frequency anomalies for the CMIP6 scenarios and for the CMIP5 RCP8.5, with respect to the historical regime frequency.

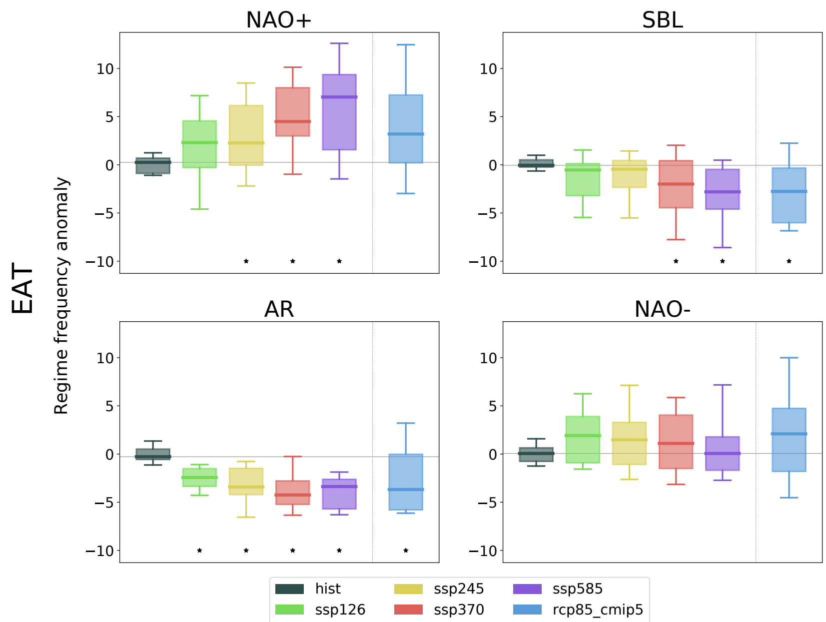

Figure 4. Whiskers plot of the multi-model distribution of WRs frequency in the CMIP6 historical (1964-2014) and future (2050-2100) reference periods, with respect to the historical regime frequency. The boxes indicate the first and third quartiles, the horizontal bar indicates the median, and the top and bottom bars indicate the 10 and 90 percentiles. The black star at the bottom indicates a significant difference at the 95% level with respect to the historical distribution using a Welch's t-test.

The results show a net increase in the frequency of the NAO+ regime in the future. For CMIP6 the signal strength increases with increased greenhouse gases concentration: there is a smaller increase for ~~SSP126~~SSP1-2.6, and a progressively larger (and statistically significant) increase for ~~SSP245, SSP370 and SSP585~~SSP2-4.5, SSP3-7.0 and SSP5-8.5. For the two most extreme scenarios the signal is robust, with the first quartiles of the 50-yr reference period ending up above the historical 90th percentile. Consistently with this picture, the trend continues till the end of the simulations for the extreme scenarios, while ~~SSP126 and SSP245~~ SSP1-2.6 and SSP2-4.5 stabilize before the end of the century (Figure 5, left panel). The net increases in the NAO+ frequency are confirmed by the positive trends over the whole 2015-2100 scenario simulations (see also Figure ~~S4~~S5, Supplementary Material). The ~~behavior~~ behaviour observed in RCP8.5 of CMIP5 is in general agreement with CMIP6, though the amplitude of the change is largely reduced with respect to ~~SSP585~~SSP5-8.5, in particular in the last part of the century, and the difference is not statistically significant with respect to the historical period.

The increase in the NAO+ frequency is accompanied by a general decrease in the AR and SBL frequency. For SBL the signal is robust for ~~SSP370 and SSP585~~SSP3-7.0 and SSP5-8.5, for which the box stands almost completely below the historical median frequency and the difference with respect to the historical is statistically significant. The two moderate scenarios also show a small decrease in the frequency, but the signal is much smaller. The AR regime (Fig 4, bottom left panel) is

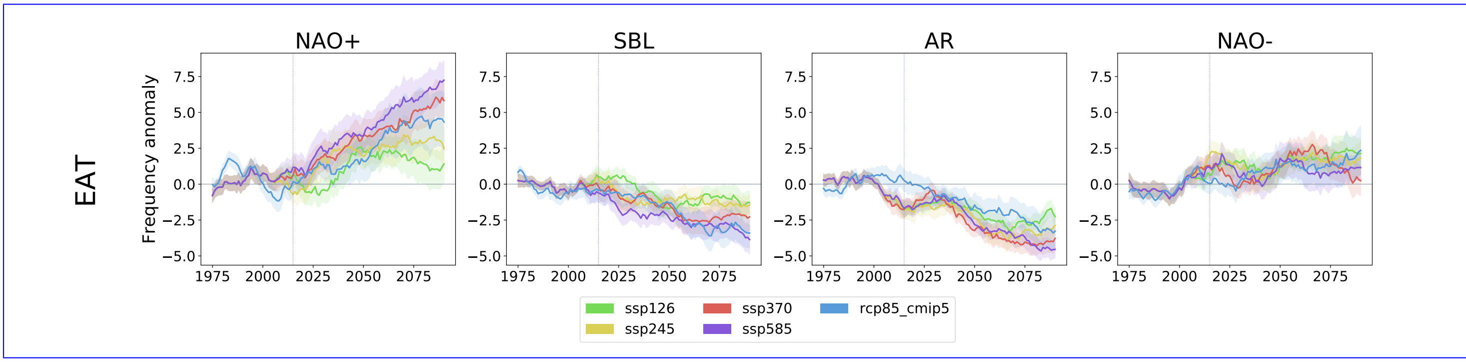

Figure 5. Ensemble means of the WRs seasonal frequency anomalies for the CMIP6 scenarios and the CMIP5 RCP8.5, with respect to the historical regime frequency. A 20-yr running mean has been applied to the time series. Shading indicates the standard error of each ensemble. A corresponding figure for the PAC sector is Figure S10 in the Supplementary Materials.

characterized by strong and statistically significant reduction in the frequency for all scenarios, as shown by the future boxes staying entirely below the historical 10th percentile. RCP8.5 shows consistent results for SBL, whose seasonal frequency projection fits remarkably well the ~~SSP585~~ SSP5-8.5 one, at least in the second half of the century (Figure 5). For AR the RCP8.5 signal is consistent but reduced in amplitude and with a larger inter-model spread with respect to both ~~SSP370 and SSP585~~SSP3-7.0 and SSP5-8.5.

The NAO- regime is characterized by a more complex response. There is a general tendency for a small increase in the regime frequency in the future, but unlike the other regimes the signal is slightly stronger in the moderate scenarios. Also, during the last 20 years the differences between the moderate and extreme scenarios amplify, showing a larger NAO- frequency for ~~SSP126 and SSP245~~ SSP1-2.6 and SSP2-4.5 and barely any change for ~~SSP370 and SSP585~~ SSP3-7.0 and SSP5-8.5 with respect to the historical period. RCP8.5 shows a consistent trend in the future and a final frequency increase larger than most CMIP6 scenarios, but still not significant with respect to the historical period. Also apparent from Figure 5 are the oscillations in the multi-model mean which appear to be quite in phase up to about 2060. ~~Further~~ These might be related to the aerosol forcing, which has been hypothesized to have driven the observed AMV oscillations (Zhang et al., 2013; Qin et al., 2020). However, it is not clear whether a similar process might be at work for the future scenario period and further analysis on this topic will be carried out in ~~future studies~~a different study.

The change in the ~~PNA~~ PAC regime frequencies is shown in Figure 6. The main changes are seen for the Pacific ~~Through~~ Trough and Bering Ridge regimes. The PT regime shows a net and statistically significant increase in frequency in all future scenarios, with some differences between them but no clear dependence on the future forcing. The RCP8.5 projected PT frequency for the future is slightly larger than the CMIP6 scenarios, but also shows a larger spread. The BR regime is projected to decrease its frequency in the CMIP6 scenarios, with a more robust decrease in the extreme ones. RCP8.5 projections show a larger reduction than the CMIP6 scenarios. The two PNA regimes show smaller and not significant variations in the future frequencies. However, whereas the PNA+ response is consistent with no change at all for all scenarios, the PNA- shows a

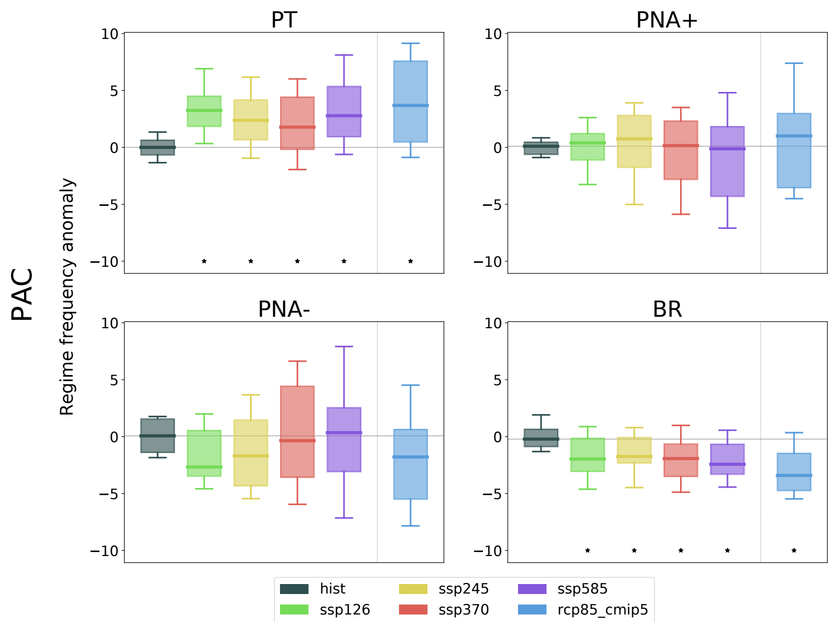

Figure 6. As Figure 4, but for ~~PNA~~ PAC regimes.

variation of the response from negative to zero change in frequency ranging from the smallest to the largest greenhouse gases concentrations. The RCP8.5 ensemble shows a reduction in the future PNA- frequency, in contrast with ~~SSP585~~SSP5-8.5. Interestingly, even if the two PNA regimes frequencies are not changing significantly, the spread in the model response increases proportionally to the CO_2 forcing. This suggests that models respond linearly to increased forcing, but that there is a large intermodel spread in the response.

3.3.2 Regime persistence

The average regime persistence also changes in the future according to CMIP6 and CMIP5 models. As shown in Figure 7 (first row), the change in the average regime duration (days) for the EAT sector is generally consistent with the direction of changes in the regime frequency. The NAO+ regime is expected to have a longer duration, with the largest increase in the average number of days per regime event up to about one day in the ~~SSP585~~ SSP5-8.5 scenario. In the two extreme scenarios, NAO+ takes the place of NAO- as the regime with longest average duration. Concurrently with the NAO+ increase, we observe a large decrease in the average duration of the AR regime and a small - but statistically significant - decrease for the SBL regime. No significant change is seen for NAO-. RCP8.5 confirms these tendencies, though the amplitude of the NAO+ and AR change is reduced with respect to CMIP6, and changes for SBL and AR are not statistically significant. In the ~~PNA~~ PAC sector (Figure 7, second row), the PT regime shows a substantial response, with a net and statistically significant increase in the average regime

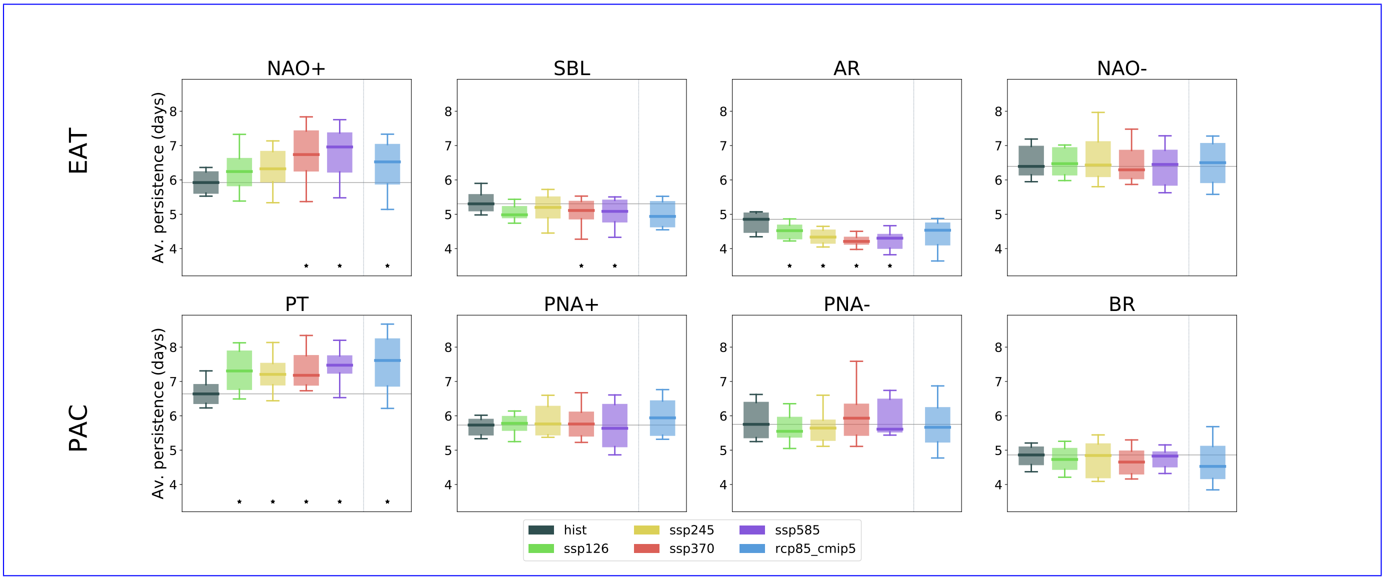

Figure 7. Average regime duration in days for the historical runs and future scenarios. Boxes are defined for the same periods as in 4, and the black stars at the bottom indicate significant variation with respect to the historical period.

duration in all future scenarios. The ~~SSP585~~ SSP5-8.5 scenario shows the largest increase, reaching an average duration of about 7.5 days in the 2050-2100 period. The response to the RCP8.5 scenario is consistent with that to ~~SSP585 , however~~ SSP5-8.5, but the model spread is larger. The other regimes do not show a clear variation of the persistence in the future in the CMIP6 ensemble, while RCP8.5 projected the BR persistence to slightly decrease in the future.

Figure 8 shows the number of regime events per 100 days. The changes in the regime frequencies might be seen as the combined effect of the changes in the regime persistence and the changes in the number of regime events. For the EAT sector, both have a comparable role in the frequency change of AR and SBL, while the increased persistence seems the main factor in the NAO+ change. For the PAC sector, the increase in PT frequency is driven by longer persistence, despite no significant change in the number of events, while the opposite is true for the change in the BR regime frequency.

4 Discussion

The projected changes in the regime frequencies in the future scenarios give a clear picture of the evolution of the variability in the mid-latitudes dynamics. For the EAT sector, the results shown in Section 3.3 are consistent with a zonalization of the mid-latitude circulation and a squeezing of the eddy-driven jet distribution around its central position (Peings et al., 2018; Oudar et al., 2020). There is generally a good correspondence between the Euro-Atlantic WRs and the North-Atlantic jet latitude index (Madonna et al., 2017; Fabiano et al., 2020). An increase in the NAO+ frequency corresponds to a more frequent central jet position, while the decrease in SBL means a lower probability of a tilted jet and a reduction in the spread of the distribution. This is also consistent with the reduced variance in terms of latitudinal shifts of the jet, observed by Barnes and Polvani (2013)

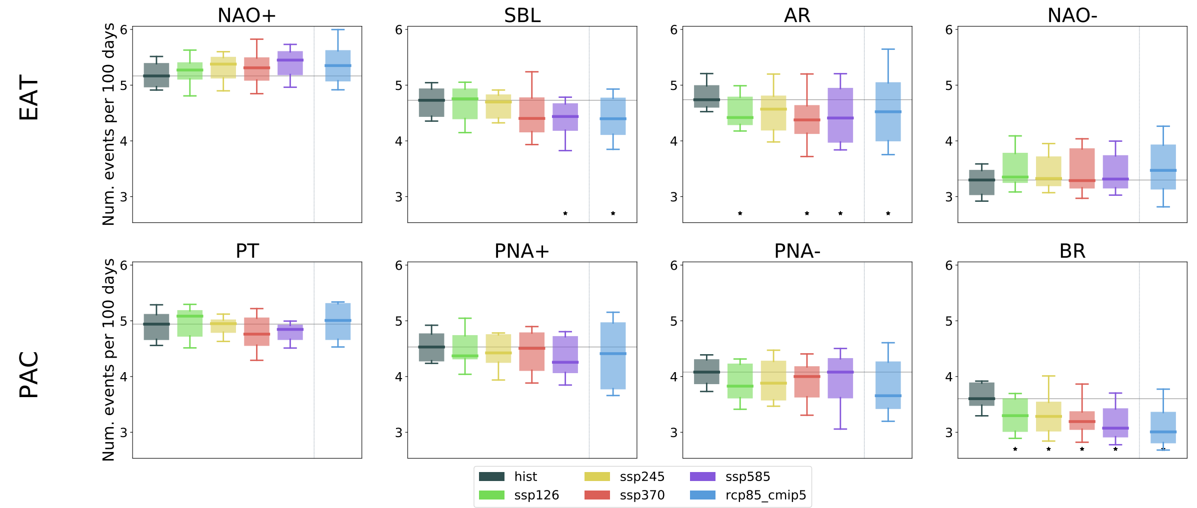

Figure 8. Number of regime events per 100 days for the historical runs and future scenarios. Boxes are defined for the same periods as in 4, and the black stars at the bottom indicate significant variation with respect to the historical period.

on the CMIP5 ensemble. Also, a less frequent AR regime would mean a reduction in the northward peak of the jet latitude distribution. The decrease in the AR and SBL frequency is in agreement with the predicted decrease in the blocking frequency over Europe, observed on CMIP5 and CMIP6 models (Davini and D'Andrea, 2020). The change of WR frequency observed for RCP8.5 is consistent with the result by Cattiaux et al. (2013a), although they observe a larger increase of NAO-, which may be due to a different treatment of the climatological mean state.

The strong increase in the NAO+ regime frequency is in line with the change of storm track activity in CMIP6 projections analysed by Harvey et al. (2020), which shows an intensification of the 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. 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. In terms of impacts, NAO+ drives mild temperatures over the Eurasian continent and a North-South precipitation dipole with increased precipitation over Northern Europe and dry conditions over South Europe. For negative NAO, colder temperatures are found in Northern Europe, and the precipitation dipole is reversed, with increased precipitation in the South (Yiou and Nogaj, 2004). The increased NAO+ frequency in the future would thus lead to higher winter precipitation in the Northern part of the continent, with a concomitant lower precipitation and higher risk of droughts over the Mediterranean region. At the same time, the increased persistence may increase the risk of flooding in Northern Europe.

With regards to the ~~PNA~~ PAC sector, the increased frequency of the PT regime and the reduction in the BR regime are consistent with the projected decrease in blocking frequency in the Bering Strait region (Davini and D'Andrea, 2020). ~~Also, the~~ The PT regime is characterized by a positive geopotential anomaly over central Canada, and its increased frequency in the future may be linked with the projected increase in the waviness index observed by Peings et al. (2017) in this region. Also, a strong decrease of the storm track activity under SSP5-8.5 has been observed over the whole Northern American continent, and an increase in the central North Pacific (Harvey et al., 2020). This agrees well with the prediction of an increased PT regime, that blocks the entrance of perturbations in the continent. In the observations, the PT regime tends to be more frequent during positive ENSO (Straus and Shukla, 2002; Weisheimer et al., 2014), therefore its increase in the future may be linked with an increased Niño-like forcing. ~~If this is the case, the fact that PT becomes more frequent in the future would suggest a higher relative importance of the tropical pacific forcing with respect to the orographic forcing in a warmer climate~~A recent study of ENSO occurrence in CMIP6 projections (Fredriksen et al., 2020) shows a tendency for an increase of ENSO variability under global warming, and interestingly this change appears to be mostly related to an increase in positive El Niño events.

4.1 Relation with changes in the mean state

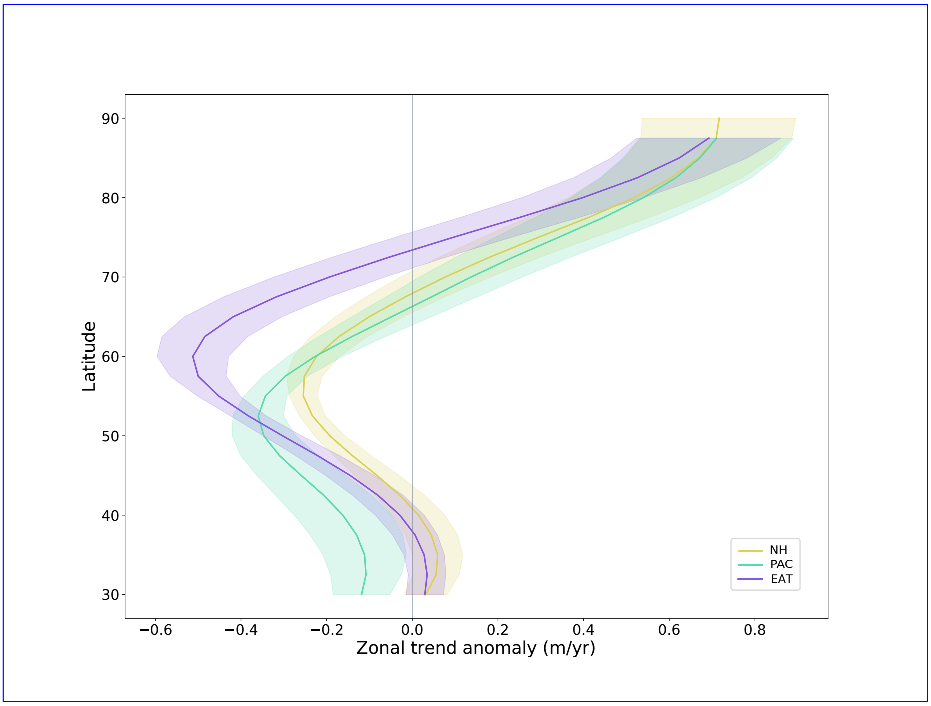

Figure 9. Multi-model average of the zonal-mean trend anomaly of the geopotential height at 500 hPa for SSP5-8.5. The zonal trend anomaly is shown for the Northern Hemisphere from 30 to 90N (NH, yellow) and for the two sectors analysed here, PAC (green) and EAT (purple), after removing the global NH trend.

The change in the regime frequency is inextricably linked to a modification of the mean geopotential height at 500 hPa. On the one hand, one can explain variations in the frequencies as the result of a global shift of the climate state towards one regime. On the other hand a change in the mean state can be interpreted as the side effect of an increase in the occurrence of some weather regimes (e.g. NAO+ and PT) and a corresponding decrease in the frequency of others (e.g. SB and BR). The change in the mean state of the geopotential height at 500 hPa during the extended boreal winter (NDJFM) for the 2015-2100 period in the SSP5-8.5 scenario multi-model ensemble simulations is analysed by taking deviations from the third-order polynomial fit of the area-weighted season-averaged Northern Hemisphere (30-90 N) time series. The projected change in the geopotential height depends on both latitude and longitude. For further insight, the multi-model mean response is split in ~~in~~ two parts:

- the zonal mean trend anomaly, shown in Figure ~~8~~9;
- the local departures from the zonal mean trend (or, equivalently, the trend of the stationary eddies), shown in Figure ~~9.~~ 10.

The multi-model average of the zonal mean trend anomaly shows a larger increase in the geopotential at high latitudes and a smaller change at mid-latitudes, peaking at about 60 N. Restricting the analysis to the EAT and ~~PNA~~ PAC sectors gives similar results, but with an intensification of the negative anomaly at mid-latitude in EAT and a southward shift of the negative peak in ~~PNA~~PAC. The trend of the stationary eddies provides further insight in the mean state change. The negative trend in the North Atlantic, west of the British Isles and south of Iceland, is consistent with a more frequent occurrence of the NAO+ regime and a decrease in the frequency of the AR regime described in Section 3.3. A positive trend over the Mediterranean region and the development of two highs, one over central Northern Canada and the other over the whole Asian continent can be noted as well. This picture is consistent with an increase of the geopotential at high latitudes and the concurrent eastward shift in phase of the stationary waves already observed in CMIP5 models (Wills et al., 2019; Simpson et al., 2014). The polar high is linked to the increased temperatures in the region, due to Arctic ~~Amplification~~amplification, and the shift in phase may be due to the decrease of the dominant zonal wavenumber of stationary waves with global warming, as found by (Wills et al., 2019).

4.2 ~~Drivers~~ Potential drivers of future circulation changes

Although future changes in regime frequencies and average persistence times are apparent in the second half of the 21st century when multi-model ensemble means are considered, a considerable spread in the model response is evident as well and may be linked to differences in the model climate feedbacks. We here ~~analyze~~ analyse possible drivers of this spread in the ~~SSP585 and rcp85~~ SSP5-8.5 and RCP8.5 scenarios. A method to investigate the model spread is to decompose the mid-latitude future response into different components related to the differential warming of the Earth's atmosphere. A set of indices ~~– also known as "remote drivers" (e.g. Zappa and Shepherd, 2017; Peings et al., 2018; Oudar et al., 2020) –~~ that (in principle) have the potential to affect ~~WR frequencies can be~~ the mid-latitude circulation and the WR frequencies are selected:

- *UTW*: the ~~Upper Tropospheric Warming~~upper tropospheric warming, computed as the temperature trend in the tropical upper troposphere (averaged between 20S and 20N and from 400 to 150 hPa);
- *AA*: the Arctic ~~Amplification~~amplification, computed as the temperature trend in the Arctic lower troposphere (averaged between 60N and 90 N and from 1000 to 700 hPa);

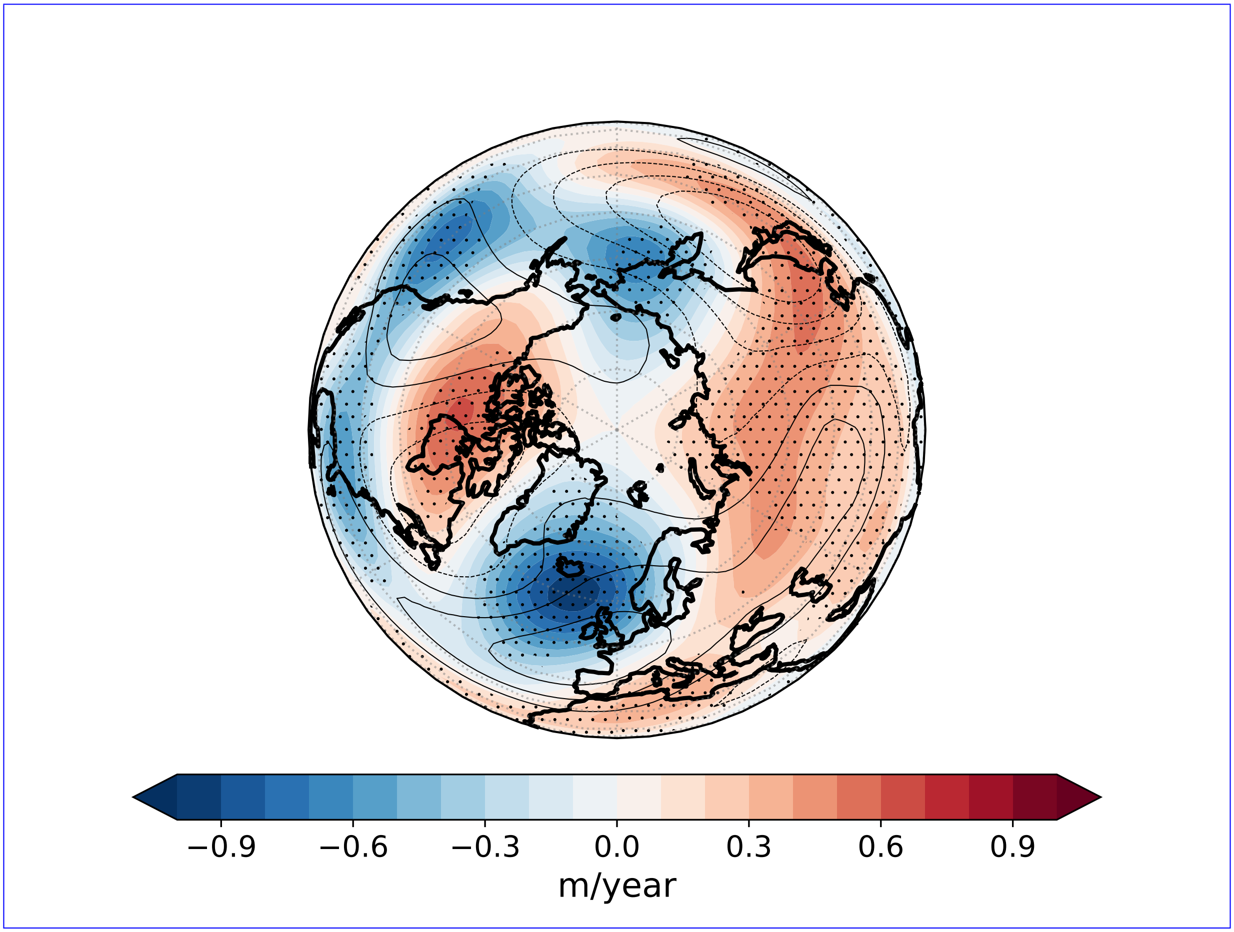

Figure 10. Multi-model average of the ~~zonal-mean trend anomaly~~ trends in the stationary eddies (shading) and multi-model mean of the ~~geopotential height at 500 hPa~~ stationary eddies for ~~SSP585~~the historical simulations (contour). Hatching indicates where the 80% of the models agree on the sign of change. The trend in the stationary eddies is ~~shown for~~ equivalent to the ~~whole hemisphere (yellow) and for~~ residual local trend, after removing the ~~two sectors analyzed here: PNA (green)~~ global and ~~EAT (purple)~~zonal components. ~~Multi-model average of the trends in the stationary eddies (shading) and multi-model mean of the stationary eddies for the historical simulations (contour). Hatching indicates where the 80% of the models agree on the sign of change.~~

– *PST*: the ~~Polar Stratospheric Temperature~~polar stratospheric temperature, i.e. the temperature trend averaged between 70 and 90 N and from 250 and 30 hPa;

– ~~*NAW*~~*TNAW*: the surface warming ~~in the~~ trend in the tropical North Atlantic (averaged between 80W and 10E, ~~20N and 75N~~10N and 30N);

– *SPGW*: the surface warming trend in the North Atlantic subpolar gyre (averaged between 60W and 20W, 40N and 70N).

The first three indices have already been considered as potential drivers of the circulation changes in future scenarios by Zappa and Shepherd (2017), Peings et al. (2017, 2018) and Oudar et al. (2020). The above-defined indices have been computed as trends over 2015-2100 (2006-2100) for SSP5-8.5 (RCP8.5). In order to ~~understand the link between these drivers~~ explore the potential links between them and the projected change in regime frequencies (Section 3.3), a multi-linear regression analysis has been performed~~, looking for the optimal sets of 2 and 3 drivers to explain the frequency trends in each sector. The optimal set is the combination of the drivers listed above with the highest R2 score (i.e. the highest explained variance).~~ . This analysis aims at finding significant relationships between the set of potential drivers and the WR frequency trends. Of course, a significant correlation between two quantities does not demonstrate the existence of a causal link - as they might be responding to a common external forcing - nor gives indication on the direction of such link. Nevertheless, this can provide an insight about the interconnection between mid-latitude climate variability and large scale changes in GCMs that deserve further investigation.

Since some of the ~~drivers~~ indices are highly correlated (see Table S3, Supplementary Material), all have been divided by the global surface warming, i.e. the global trend of the atmospheric surface temperature (tas) during 2015-2100. After this operation, moderate correlations remain between AA and ~~NAW (0.42)and UTW and NAW (-0.41~~TNAW (0.28), TNAW and SPGW (0.28), SPGW and UTW (-0.25), while other ~~cross-correlations~~ correlations have absolute values below 0.2 (see Supplementary Materials, Table ~~S3~~S2). All ~~drivers are standardized to unit mean and~~ indices have been standardized to zero mean and unit variance before performing the regressions. The frequency trends were also divided by the global warming. ~~The results are shown in~~

Figure 10 ~~, where~~ shows the optimal sets of 2 and 3 indices to fit the regime frequency trends in each sector: the columns represent the different regimes and the rows the ~~optimal sets of drivers for the two sectors.~~

~~Regression coefficients for the best sets of 2 and 3 drivers, and the two sectors. The indices for the "remote drivers" have been standardized before performing the multi-linear regression. The regression coefficients are shown in color code and the respective statistical significance is indicated by the white circles (big circle: p < 0.01; small circle: p < 0.05). The mean and standard deviation of the drivers before standardization are shown in the Supplementary Material, Table S3.~~

indices. The optimal set is the combination of the potential drivers listed above with the highest R2 score (i.e. the highest explained variance). The regression coefficients are shown in color code and the respective statistical significance is indicated by the white circles (big circle: $p < 0.01$; small circle: $p < 0.05$). However, the effective size of the sample may be smaller than the 38 models considered here (19 for both SSP5-8.5 and RCP8.5), since some models are closely related to each other and this might lower the significance of the regressions. The overall R2 score of the regressions are 0.28 (~~0.31) and 0.31 (~~0.21) and 0.34 (0.30) for the 2 and 3 ~~drivers~~ indices set of the EAT (~~PNA~~PAC) sector. However, the score varies strongly among the regimes and it is higher for NAO+ (0.5) and NAO- (~~0.6~~0.55) and very low for SBL and AR (0.1-0.2). On ~~PNA~~ PAC the score is more uniform, between ~~0.3~~ 0.25 and 0.4 for all regimes (see Figure ~~S7).~~ S12). The regression model with all 5 indices is shown in Figure S13.

For the EAT regimes, the dominant ~~drivers of the intermodel spread appear to be the Polar Stratospheric Temperature~~ connections are found with the polar stratospheric temperature and the Arctic ~~Amplification, while UTW has a secondary~~

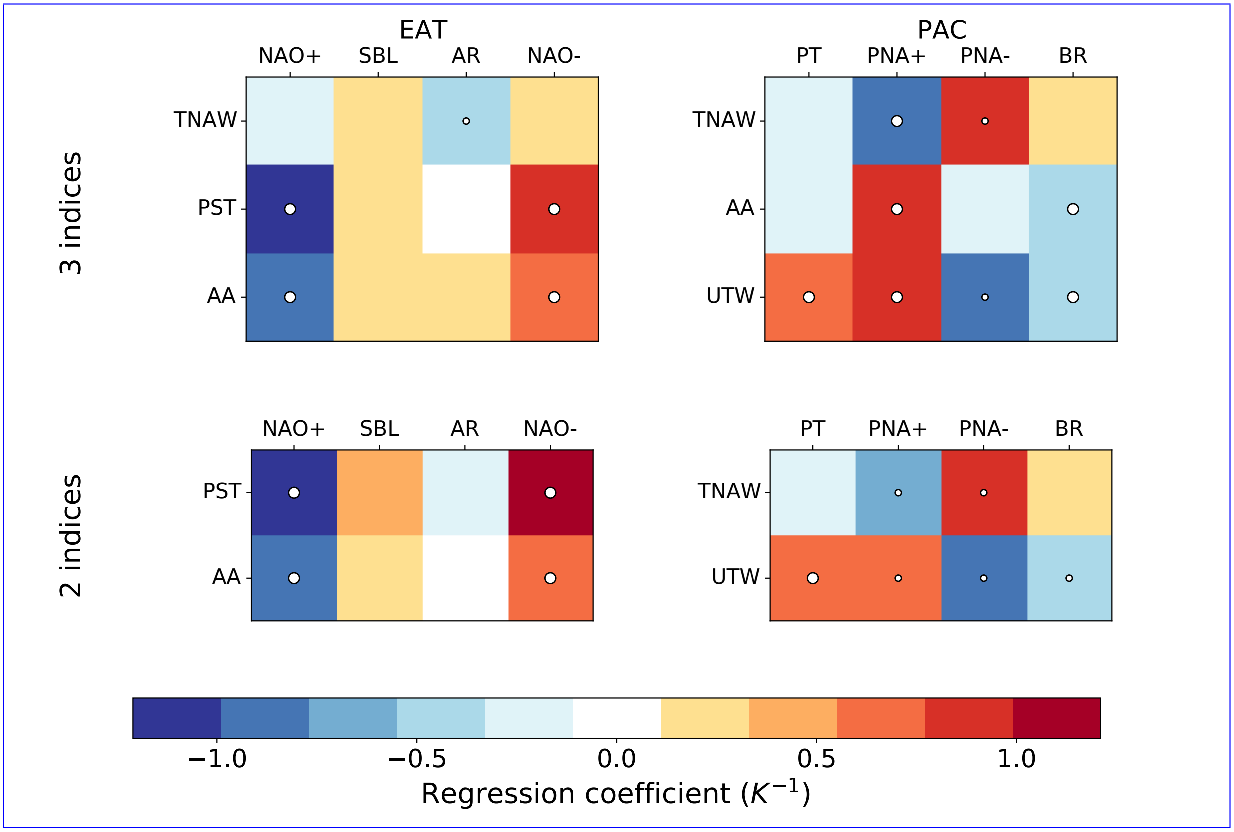

Figure 11. Regression coefficients for the best sets of 2 and 3 indices to fit the projected trends of the weather regimes frequencies in the two sectors. The indices have been standardized before performing the multi-linear regression. The regression coefficients are shown in color code and the respective statistical significance is indicated by the white circles (big circle: $p < 0.01$; small circle: $p < 0.05$). The standard deviation of the indices before standardization are shown in the Supplementary Material, Table S1.

~~role and NAW gives only a marginal contribution (not shown). The Polar Stratospheric Temperature~~ amplification. The polar stratospheric temperature explains a considerable fraction of the spread in the EAT response, with warmer temperatures driving a decrease in NAO+ states and an increase in NAO-. Indeed, a warm polar stratosphere is linked to a decrease in the ~~Polar Vortex Strength~~polar vortex strength, and a negative NAO index is more commonly observed in response to a weak polar vortex ~~Baldwin and Dunkerton (2001)~~(Baldwin and Dunkerton, 2001; Ambaum and Hoskins, 2002). This is in line with other works that found a significant relation of the PST with the North Atlantic circulation changes (Manzini et al., 2014; Zappa and Shepherd, 2017; Peings et al., 2017). The Arctic ~~Amplification~~ amplification goes in the same direction as PST and ~~tends to reduce~~ is linked with a reduction in the NAO+ and ~~increase~~ an increase in the NAO- frequency, which is consistent with the expected contribution of AA towards a weakening and equatorward shift of the jet (Barnes and Screen, 2015; Peings et al., 2018; Cohen et al., 2019). The other indices only explain a small portion of the remaining variance. The significant negative correlations of TNAW with AR and of SPGW with NAO+ (Figure S13) may indicate the influence of local SSTs on those regimes, although the NAO+ frequency change may also have an impact on the local surface temperature and SPGW. The role of the UTW is less clear, since the only significant positive correlation is with NAO- (Figure S13), but a stronger meridional gradient in the upper troposphere would instead be expected to push towards the zonalization of the jet ~~in the EAT sector. Over the PNA sector, a considerable~~ (Barnes and Screen, 2015; Peings et al., 2018).

Over the PAC sector, the most significant relations are found with UTW and TNAW, but the 2-indices model explains a smaller fraction of the ~~variance is explained by the relative warming of North Atlantic (NAW) and by the Arctic Amplification. The other drivers - PST and UTW - all give marginal contributions. The AA~~ total variance than for EAT. The UTW is positively correlated with PT and PNA+, and negatively with ~~BR. The NAW~~ PNA- and BR, which resembles the influence of positive ENSO on the PAC regimes (Straus and Shukla, 2002). The TNAW has a significant regression ~~with all regimes, being positively correlated with PNA- and BR and negatively~~ coefficient with the two PNA regimes, opposite to that of the UTW, and only small non-significant regressions with PT and ~~PNA+~~BR. This influence of a warmer tropical North Atlantic on the Pacific regimes may be due to ~~a more frequent negative ENSO, which tends to increase the PNA- and BR and to decrease the PT and PNA+ regimes occurrence (Straus and Shukla, 2002)~~its link with negative ENSO. In fact, there are indications that the strength of the Pacific Walker circulation may be controlled by the Atlantic ocean temperature (McGregor et al., 2014; Li et al., 2016), with a warmer Atlantic linked to more frequent La Niña-like conditions. However, this significant regression may also show the response to a common external forcing, influencing both the TNAW and the PNA regimes, like a change in the ENSO forcing (Fredriksen et al., 2020). Among the other indices, the AA is linked with an increase in PNA+ and a decrease in BR, while PST and SPGW only have significant positive correlations with PT and BR respectively (Figure S13).

5 Conclusions

We proposed here an alternative view of future changes in the atmospheric circulation at Northern mid-latitudes, considering the future trends in ~~Weather Regimes~~ weather regimes frequency and persistence over the Euro-Atlantic and Pacific-North American sectors, as projected by CMIP5 and CMIP6 models. The CMIP6 ensemble shows a non negligible improvement in the reproduction of the ~~Weather Regimes~~ weather regimes when compared to CMIP5 (Section 3.2). A better simulation of the regime patterns is in particular evident over the EAT sector, however both sectors show an improvement in the other metrics considered (i.e. the variance ratio and the regime frequency bias, see Section 3.2). The model biases in simulating the observed regime centroids, frequencies and variance ratio are known and documented in literature ~~(Dawson et al., 2012; Strommen et al., 2019; Fabiano et al., 2020; Weisheimer et al., 2014)~~(Dawson et al., 2012; Weisheimer et al., 2014;
. The improvements of CMIP6 models compared to CMIP5 in this respect are therefore encouraging.

Over the EAT sector an increase in the NAO+ frequency and persistence during the second half of the 21st century is observed in all scenarios, with larger changes in the ~~SSP370 and SSP585~~ SSP3-7.0 and SSP5-8.5 (Section 3.3.1). This increase is accompanied by a decrease in the AR frequency (and persistence) in all scenarios and a decrease in the SBL frequency, more pronounced in the most extreme scenarios. The NAO- regime shows a small positive trend in all scenarios. These trends are consistent with changes in the mean geopotential height state, that shows an increase at high latitudes and a pronounced eastward shift of the stationary eddies (Section 4.1). A significant fraction of the spread of the model response over the EAT sector is ~~explained by the Polar Stratospheric Temperature~~ related with the spread in the polar stratospheric temperature and the Arctic ~~Amplification~~ amplification in future projections (Section 4.2). The increase of the NAO+ regime is consistent with a squeezing of the jet around the central position (Peings et al., 2018; Oudar et al., 2020) and with a reduced meridional variability of the jet (Barnes and Polvani, 2013). ~~In the PNA~~

In the PAC sector the future trends are characterized by an increase in the PT regime occurrence, with a concomitant decrease in the BR regime frequency. The two PNA regimes do not show clear trends in the future scenarios. The ~~main drivers of the intermodel spread are~~ intermodel spread in the PAC WR trends correlates significantly with the upper tropospheric warming in the tropics and the warming of the ~~North Atlantic and the Arctic Amplification~~ tropical North Atlantic (Section 4.2). The increase in the PT regime frequency indicates a larger relative importance of tropical forcing versus orographic forcing in perturbing the mean flow. The decrease in the BR regime is consistent with changes in the mean state (Section 4.1) and with a decrease in the blocking frequency in the Bering Strait (Davini and D'Andrea, 2020).

The regime perspective presented in this work provides a clear picture of future changes in the wintertime mid-latitude atmospheric circulation and also introduces a suitable framework to study the impact of extreme weather in the future scenarios. The projected change in the regime frequencies are associated with important changes in the temperature and precipitation statistics over different regions. For example an increase in frequency of a strong zonal flow regime (i.e. NAO+) over the Atlantic can lead to an enhanced flood risk due to its connection with stormy weather over northwestern Europe and the British Isles (Yiou and Nogaj, 2004). In this respect it is worth noting that the extreme rainfall in the UK during winter 2013-2014 resulted from this type of atmospheric circulation (Knight et al., 2017) and human-induced climate change was recognised as the major driver of such an extreme (Vautard et al., 2016). Also, the Mediterranean region might suffer from ~~summetime~~ summertime dry spells and heat waves in response to a deficit in precipitation during winter (Vautard et al., 2007), which is more likely to occur with increased NAO+ frequency.

Code availability. The WRtool package is freely available at https://github.com/fedef17/WRtool.

Author contributions. FF conducted most of the data analyses and all visualisations and wrote the paper. VM performed part of the data analysis. VM, PD, PG and SC all commented, organized and wrote parts of the paper.

Competing interests. The authors declare that they have no conflict of interests.

Acknowledgements. The authors acknowledge support by the PRIMAVERA project of the Horizon 2020 Research Programme, funded by the European Commission under Grant Agreement 641727.

References

Ambaum, M. H. and Hoskins, B. J.: The NAO troposphere–stratosphere connection, Journal of Climate, 15, 1969–1978, 2002.

Athanasiadis, P. J., Wallace, J. M., and Wettstein, J. J.: Patterns of wintertime jet stream variability and their relation to the storm tracks, Journal of the Atmospheric Sciences, 67, 1361–1381, 2010.

Baldwin, M. P. and Dunkerton, T. J.: Stratospheric harbingers of anomalous weather regimes, Science, 294, 581–584, 2001.

Barnes, E. A. and Polvani, L.: Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the CMIP5 models, Journal of Climate, 26, 7117–7135, 2013.

Barnes, E. A. and Polvani, L. M.: CMIP5 projections of Arctic amplification, of the North American/North Atlantic circulation, and of their relationship, Journal of Climate, 28, 5254–5271, 2015.

Barnes, E. A. and Screen, J. A.: The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it?, Wiley Interdisciplinary Reviews: Climate Change, 6, 277–286, 2015.

Barnston, A. G. and Livezey, R. E.: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, Monthly weather review, 115, 1083–1126, 1987.

Beerli, R. and Grams, C. M.: Stratospheric modulation of the large-scale circulation in the Atlantic–European region and its implications for surface weather events, Quarterly Journal of the Royal Meteorological Society, 145, 3732–3750, 2019.

Blackport, R. and Screen, J. A.: Insignificant effect of Arctic amplification on the amplitude of midlatitude atmospheric waves, Science advances, 6, eaay2880, 2020.

Brunner, L., Schaller, N., Anstey, J., Sillmann, J., and Steiner, A. K.: Dependence of present and future European temperature extremes on the location of atmospheric blocking, Geophysical research letters, 45, 6311–6320, 2018.

Casola, J. H. and Wallace, J. M.: Identifying weather regimes in the wintertime 500-hPa geopotential height field for the Pacific–North American sector using a limited-contour clustering technique, Journal of applied meteorology and climatology, 46, 1619–1630, 2007.

Cassou, C.: Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic Oscillation, Nature, 455, 523, 2008.

Cattiaux, J., Douville, H., and Peings, Y.: European temperatures in CMIP5: origins of present-day biases and future uncertainties, Climate dynamics, 41, 2889–2907, 2013a.

Cattiaux, J., Quesada, B., Arakélian, A., Codron, F., Vautard, R., and Yiou, P.: North-Atlantic dynamics and European temperature extremes in the IPSL model: sensitivity to atmospheric resolution, Climate dynamics, 40, 2293–2310, 2013b.

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., Ballinger, T., Bhatt, U., Chen, H., Coumou, D., et al.: Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather, Nature Climate Change, pp. 1–10, 2019.

Corti, S., Molteni, F., and Palmer, T.: Signature of recent climate change in frequencies of natural atmospheric circulation regimes, Nature, 398, 799, 1999.

Davini, P. and D'Andrea, F.: From CMIP3 to CMIP6: Northern Hemisphere atmospheric blocking simulation in present and future climate, Journal of Climate, 33, 10 021–10 038, 2020.

Dawson, A. and Palmer, T.: Simulating weather regimes: Impact of model resolution and stochastic parameterization, Climate Dynamics, 44, 2177–2193, 2015.

Dawson, A., Palmer, T., and Corti, S.: Simulating regime structures in weather and climate prediction models, Geophysical Research Letters, 39, 2012.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

Di Capua, G. and Coumou, D.: Changes in meandering of the Northern Hemisphere circulation, Environmental Research Letters, 11, 094 028, 2016.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, 2016.

Fabiano, F., Christensen, H., Strommen, K., Athanasiadis, P., Baker, A., Schiemann, R., and Corti, S.: Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance, Climate Dynamics, 54, 2020.

Fredriksen, H.-B., Berner, J., Subramanian, A. C., and Capotondi, A.: How Does El Niño–Southern Oscillation Change Under Global Warming—A First Look at CMIP6, Geophysical Research Letters, 47, e2020GL090 640, 2020.

Hannachi, A., Straus, D. M., Franzke, C. L., Corti, S., and Woollings, T.: Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere, Reviews of Geophysics, 55, 199–234, 2017.

Harvey, B., Cook, P., Shaffrey, L., and Schiemann, R.: The Response of the Northern Hemisphere Storm Tracks and Jet Streams to Climate Change in the CMIP3, CMIP5, and CMIP6 Climate Models, Journal of Geophysical Research: Atmospheres, 125, e2020JD032 701, 2020.

Hassanzadeh, P., Kuang, Z., and Farrell, B. F.: Responses of midlatitude blocks and wave amplitude to changes in the meridional temperature gradient in an idealized dry GCM, Geophysical Research Letters, 41, 5223–5232, 2014.

Hoskins, B. and Woollings, T.: Persistent extratropical regimes and climate extremes, Current Climate Change Reports, 1, 115–124, 2015.

Kimoto, M. and Ghil, M.: Multiple flow regimes in the Northern Hemisphere winter. Part II: Sectorial regimes and preferred transitions, Journal of the atmospheric sciences, 50, 2645–2673, 1993.

Knight, J. R., Maidens, A., Watson, P. A., Andrews, M., Belcher, S., Brunet, G., Fereday, D., Folland, C. K., Scaife, A. A., and Slingo, J.: Global meteorological influences on the record UK rainfall of winter 2013–14, Environmental Research Letters, 12, 074 001, 2017.

Li, X., Xie, S.-P., Gille, S. T., and Yoo, C.: Atlantic-induced pan-tropical climate change over the past three decades, Nature Climate Change, 6, 275–279, 2016.

Loikith, P. C. and Broccoli, A. J.: The Influence of Recurrent Modes of Climate Variability on the Occurrence of Winter and Summer Extreme Temperatures over North America, Journal of Climate, 27, 1600–1618, https://doi.org/10.1175/JCLI-D-13-00068.1, https://doi.org/10.1175/JCLI-D-13-00068.1, 2014.

Ma, L., Woollings, T., Williams, R. G., Smith, D., and Dunstone, N.: How Does the Winter Jet Stream Affect Surface Temperature, Heat Flux, and Sea Ice in the North Atlantic?, Journal of Climate, 33, 3711–3730, 2020.

Madonna, E., Li, C., Grams, C. M., and Woollings, T.: The link between eddy-driven jet variability and weather regimes in the North Atlantic-European sector, Quarterly Journal of the Royal Meteorological Society, 143, 2960–2972, 2017.

Manzini, E., Karpechko, A. Y., Anstey, J., Baldwin, M., Black, R., Cagnazzo, C., Calvo, N., Charlton-Perez, A., Christiansen, B., Davini, P., et al.: Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere-troposphere coupling, Journal of Geophysical Research: Atmospheres, 119, 7979–7998, 2014.

McGregor, S., Timmermann, A., Stuecker, M. F., England, M. H., Merrifield, M., Jin, F.-F., and Chikamoto, Y.: Recent Walker circulation strengthening and Pacific cooling amplified by Atlantic warming, Nature Climate Change, 4, 888–892, 2014.

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., et al.: The SSP greenhouse gas concentrations and their extensions to 2500, Geoscientific Model Development Discussions, pp. 1–77, 2019.

Michelangeli, P.-A., Vautard, R., and Legras, B.: Weather regimes: Recurrence and quasi stationarity, Journal of the atmospheric sciences, 52, 1237–1256, 1995.

Molteni, F., Roberts, C. D., Senan, R., Keeley, S. P., Bellucci, A., Corti, S., Franco, R. F., Haarsma, R., Levine, X., Putrasahan, D., Roberts, M. J., and Terray, L.: Boreal-winter teleconnections with tropical Indo-Pacific rainfall in HighResMIP historical simulations from the PRIMAVERA project, Climate Dynamics, pp. 1–31, 2020.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., et al.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, 2010.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, https://gmd.copernicus.org/articles/9/3461/2016/, 2016.

Oudar, T., Cattiaux, J., and Douville, H.: Drivers of the Northern Extratropical Eddy-Driven Jet Change in CMIP5 and CMIP6 Models, Geophysical Research Letters, 47, e2019GL086 695, 2020.

Overland, J. E., Dethloff, K., Francis, J. A., Hall, R. J., Hanna, E., Kim, S.-J., Screen, J. A., Shepherd, T. G., and Vihma, T.: Nonlinear response of mid-latitude weather to the changing Arctic, Nature Climate Change, 6, 992–999, 2016.

Palmer, T. N.: A nonlinear dynamical perspective on climate prediction, Journal of Climate, 12, 575–591, 1999.

Peings, Y., Cattiaux, J., Vavrus, S., and Magnusdottir, G.: Late twenty-first-century changes in the midlatitude atmospheric circulation in the CESM large ensemble, Journal of Climate, 30, 5943–5960, 2017.

Peings, Y., Cattiaux, J., Vavrus, S. J., and Magnusdottir, G.: Projected squeezing of the wintertime North-Atlantic jet, Environmental Research Letters, 13, 074 016, 2018.

Pithan, F. and Mauritsen, T.: Arctic amplification dominated by temperature feedbacks in contemporary climate models, Nature Geoscience, 7, 181–184, 2014.

Qin, M., Dai, A., and Hua, W.: Aerosol-forced multidecadal variations across all ocean basins in models and observations since 1920, Science advances, 6, eabb0425, 2020.

Renwick, J. A. and Wallace, J. M.: Relationships between North Pacific wintertime blocking, El Niño, and the PNA pattern, Monthly Weather Review, 124, 2071–2076, 1996.

Robertson, A. W. and Ghil, M.: Large-scale weather regimes and local climate over the western United States, Journal of Climate, 12, 1796–1813, 1999.

Schaller, N., Sillmann, J., Anstey, J., Fischer, E. M., Grams, C. M., and Russo, S.: Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles, Environmental Research Letters, 13, 054 015, 2018.

Screen, J. A. and Simmonds, I.: The central role of diminishing sea ice in recent Arctic temperature amplification, Nature, 464, 1334–1337, 2010.

Screen, J. A. and Simmonds, I.: Amplified mid-latitude planetary waves favour particular regional weather extremes, Nature Climate Change, 4, 704–709, 2014.

Shaw, T., Baldwin, M., Barnes, E. A., Caballero, R., Garfinkel, C., Hwang, Y.-T., Li, C., O'Gorman, P., Rivière, G., Simpson, I., et al.: Storm track processes and the opposing influences of climate change, Nature Geoscience, 9, 656–664, 2016.

Simpson, I. R., Shaw, T. A., and Seager, R.: A Diagnosis of the Seasonally and Longitudinally Varying Midlatitude Circulation Response to Global Warming, Journal of the Atmospheric Sciences, 71, 2489–2515, https://doi.org/10.1175/JAS-D-13-0325.1, https://doi.org/10.1175/JAS-D-13-0325.1, 2014.

Smyth, P., Ide, K., and Ghil, M.: Multiple regimes in northern hemisphere height fields via mixturemodel clustering, Journal of the Atmospheric Sciences, 56, 3704–3723, 1999.

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M., and Santos, J. A.: European temperature responses to blocking and ridge regional patterns, Climate Dynamics, 50, 457–477, 2018.

Straus, D. M. and Shukla, J.: Does ENSO force the PNA?, Journal of climate, 15, 2340–2358, 2002.

Straus, D. M., Corti, S., and Molteni, F.: Circulation regimes: Chaotic variability versus SST-forced predictability, Journal of climate, 20, 2251–2272, 2007.

Strommen, K., Mavilia, I., Corti, S., Matsueda, M., Davini, P., von Hadenberg, J., Vidale, P., and Mizuta, R.: The Sensitivity of Euro-Atlantic Regimes to Model Horizontal Resolution, arXiv preprint arXiv:1905.07046, 2019.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, 2001.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, 2012.

Ullmann, A., Fontaine, B., and Roucou, P.: Euro-Atlantic weather regimes and Mediterranean rainfall patterns: present-day variability and expected changes under CMIP5 projections, International journal of climatology, 34, 2634–2650, 2014.

Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131, 2961–3012, 2005.

Vautard, R., Yiou, P., D'Andrea, F., de Noblet, N., Viovy, N., Cassou, C., Polcher, J., Ciais, P., Kageyama, M., and Fan, Y.: Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit, Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL028001, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL028001, 2007.

Vautard, R., Yiou, P., Otto, F., Stott, P., Christidis, N., Van Oldenborgh, G. J., and Schaller, N.: Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events, Environmental Research Letters, 11, 114 009, 2016.

Vavrus, S. J., Wang, F., Martin, J. E., Francis, J. A., Peings, Y., and Cattiaux, J.: Changes in North American atmospheric circulation and extreme weather: Influence of Arctic amplification and Northern Hemisphere snow cover, Journal of Climate, 30, 4317–4333, 2017.

Wallace, J. M. and Gutzler, D. S.: Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter, Monthly Weather Review, 109, 784–812, https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2, https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2, 1981.

Weisheimer, A., Corti, S., Palmer, T., and Vitart, F.: Addressing model error through atmospheric stochastic physical parametrizations: Impact on the coupled ECMWF seasonal forecasting system, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372, 20130 290, 2014.

Wills, R. C., White, R. H., and Levine, X. J.: Northern Hemisphere stationary waves in a changing climate, Current climate change reports, 5, 372–389, 2019.

Woollings, T., Hannachi, A., and Hoskins, B.: Variability of the North Atlantic eddy-driven jet stream, Quarterly Journal of the Royal Meteorological Society, 136, 856–868, 2010.

Woollings, T., Barriopedro, D., Methven, J., Son, S.-W., Martius, O., Harvey, B., Sillmann, J., Lupo, A. R., and Seneviratne, S.: Blocking and its response to climate change, Current climate change reports, 4, 287–300, 2018.

Yiou, P. and Nogaj, M.: Extreme climatic events and weather regimes over the North Atlantic: When and where?, Geophysical Research Letters, 31, 2004.

Zappa, G. and Shepherd, T. G.: Storylines of atmospheric circulation change for European regional climate impact assessment, Journal of Climate, 30, 6561–6577, 2017.

Zappa, G., Hoskins, B. J., and Shepherd, T. G.: Improving climate change detection through optimal seasonal averaging: The case of the North Atlantic jet and European precipitation, Journal of Climate, 28, 6381–6397, 2015a.

Zappa, G., Hoskins, B. J., and Shepherd, T. G.: The dependence of wintertime Mediterranean precipitation on the atmospheric circulation response to climate change, Environmental Research Letters, 10, 104 012, 2015b.

Zhang, R., Delworth, T. L., Sutton, R., Hodson, D. L., Dixon, K. W., Held, I. M., Kushnir, Y., Marshall, J., Ming, Y., Msadek, R., et al.: Have aerosols caused the observed Atlantic multidecadal variability?, Journal of the Atmospheric Sciences, 70, 1135–1144, 2013.