Answers to the reviewers

egusphere-2025-1373

"Summertime Arctic and North Atlantic-Eurasian Circulation Regimes under Climate Change"

By Johannes Müller, Oskar Landgren, and Dörthe Handorf

We are very grateful to the reviewers for their insightful feedback and constructive suggestions. Their suggestions have been highly valuable for refining our manuscript, and we have carefully considered and addressed each point below. As various issues were mentioned by both reviewers, we have provided the answers to their comments in one document.

In the document the reviewers' comments are written in black, and our responses in green.

Please note the following general comment, which affects the whole document: For the Arctic regimes, the East-West Dipole regime has been renamed to Cluster 5 due to the low similarity between the patterns obtained using the K-Means and SANDRA clustering methods after recalculating. Please, refer to Figure R15 and R16 for details.

Reviewer 1:

This study looks at the change of circulation regimes under climate change for an Arctic and Atlantic-Eurasian domain. It finds that the NAO+/AO- regimes will occur more often in the future, while the NAO-/AO+ regimes will occur less.
Overall, the study is a valuable addition to the scientific literature on weather regimes, presenting important results for the Arctic. However, the description and discussion of the methods and results needs to be substantially improved before I would find this ready for publication.
I have put comments in the pdf and summarise my main points here.

My major concerns are:
1. The authors use pseudo-PC's to identify and assign regimes for the CMIP6 model data. While I think this is the most suitable approach for studying changes in regime occurrence, I question whether it is also suitable to study changes in the regime patterns. By projecting the CMIP6 data onto the ERA5 PC's you introduce a bias of sorts and I wonder how different the results would be when using the model's own PC's?

In our study, we follow the projection method introduced e.g., in Fabiano et al. (2020) and Fabiano et al. (2021). Projection ensures that all calculations are done in the same reference space and allows comparison of regime patterns in a consistent way. On the other hand (see also comment 1 from reviewer 2) the projection method requires that the model sub-space is contained in the reference space. We agree that this has to be proven. To do this, we follow the suggestion of reviewer 2, and use the "quantization error" proposed by Quagraine et al. (2020) to quantify, if the model sub-space is contained in the reference space.

The "quantization error" is defined as the mean error of each day's pattern with respect to the reference cluster pattern to which it belongs (Quagraine et al., 2020). As explained in Quagraine et al. (2020), by comparing the error mapping of the GCMs to that of the ERA5 data, one can assess whether the GCM patterns have a greater error than the reanalysis. In this case, the GCM sub-space is not contained in the reference space spanned by the reanalysis data.

Quantization Error per Cluster and Model, ATL KME

Model	SCAN	ATL	NAO +	NAO -	DIPOLE	Overall
ACCESS-CM2	575.672	583.456	621.001	606.375	600.649	597.395
ACCESS-ESM1	585.837	583.244	620.655	624.421	611.495	604.832
CESM2-WACCM	645.859	642.985	653.474	666.521	664.591	654.298
CNRM-CM6-1	607.525	629.074	639.457	644.464	633.262	630.517
CNRM-CM6-1-HR	642.063	671.858	690.407	699.495	667.168	674.397
CNRM-ESM2-1	611.466	619.264	633.677	647.915	626.018	627.650
GFDL-CM4	587.294	603.810	621.910	634.554	613.619	611.404
GFDL-ESM4	589.004	602.507	611.771	614.142	610.765	605.276
HadGEM3-GC31-LL	574.141	581.229	609.238	605.037	581.323	590.022
INM-CM4-8	518.897	522.371	537.099	537.321	525.698	527.943
INM-CM5-0	507.084	523.087	533.377	531.791	523.652	522.922
IPSL-CM6A-LR	614.781	624.652	637.107	639.145	626.827	628.562
KACE-1-0-G	577.886	585.065	613.482	614.743	586.711	595.784
MIROC-ES2L	467.766	462.031	478.652	485.639	477.913	473.847
MIROC6	514.705	537.282	543.912	548.557	543.060	536.177
MPI-ESM1-2-HR	621.226	621.676	643.672	640.489	635.953	632.574
MPI-ESM1-2-LR	605.721	616.337	628.612	630.176	624.012	620.294
MRI-ESM2-0	653.606	658.595	702.223	694.701	672.542	676.687
NESM3	625.442	647.071	660.881	661.795	634.294	645.425
NorESM2-MM	640.100	638.041	670.692	690.637	653.364	659.306
UKESM1-0-LL	578.422	593.884	603.218	603.929	597.229	594.869
ERA5	664.014	679.315	680.190	685.441	683.051	677.719

Figure R1: Quantization Error per regime and model, North-Atlantic-Eurasian region, KME method, historical period

Figure R2: As Figure R2, but for the Arctic region.

Fig. R1 shows the quantization error for the North-Atlantic-Eurasian regimes for ERA5 (last row) and the historical simulations from 1985-2014, obtained with the K-Means (KME) clustering method. Except for NAO+ respectively NAO- for only 2 respectively 3 models the GCM patterns have smaller errors than the ERA5 daily patterns, which justifies the projection approach. A similar conclusion can be drawn for the Arctic regimes, see Fig. R2. The majority of the GCMs have smaller errors than the ERA5 daily patterns, except 3 models for the AO+ regime, 3 models for the AO- regime, and one model for the East-West dipole regime. We will include that justification into the revised manuscript's appendix.

To underline the applicability of the projection approach for the calculation of the spatial regime structure (that is the simulated circulation regime framework in our terminology, Fig. 1 in the manuscript), we here show a comparison of joint regime patterns for the North-Atlantic Eurasian regimes for the extended summer season for the CMIP6 historical simulations 1985-2014.

In Fig. R3, upper row, we show the common simulated circulation regimes for the CMIP6 historical simulations 1985-2014. These are obtained by projecting the common climate model data onto the ten-dimensional ERA5 reference state space, and K-Means clustering of the resulting Pseudo-PCs. Fig. R3, lower row, displays the common circulation regimes, directly obtained by performing a common EOF analysis for the CMIP6 historical simulations 1985-2014 and subsequent K-Means clustering of the PCs. It is clearly visible that both approaches reveal similar patterns.

Please note that an improved explanation on the common simulated circulation regime framework will be given in the revised text for Method section, L122-L138 (compare answer to reviewer 2, comment 4). We will change the text from L122-L138 to:

"The common simulated circulation regime framework enables the possibility to compare the spatial structure between reanalysis and the entire CMIP6 model ensemble. First the preprocessed data of each climate model are merged into a single data file along the temporal axis, either for the historical period 1985-2014 or the future period 2070-2099. Instead of performing a common EOF analysis and obtaining common PCs, (as described e.g. in Benestad et al., 2023), here the common climate model data are projected onto the ten-dimensional reference state space, determined from the ERA5 data, resulting in ten common Pseudo-PCs. The common Pseudo-PCs serve as input data for the K-Means clustering algorithm. Five common simulated circulation regimes are obtained for each time period, representing the joint regimes for the entire model ensemble."

Figure R3: North-Atlantic Eurasian Circulation regime patterns in terms of SLP anomalies for MJJASO, for CMIP6 models in the historical period 1985-2014 obtained with K-Means clustering. Upper row: Common simulated circulation regime approach, with model data projected onto the ERA5 state space, lower row: Common regime approach with state space spanned by common EOFs from climate model data.

2.Also the other choices made in the method section are not in all case thoroughly argued for, for example the use of 10 PC's and not e.g. 15. I believe the paper would benefit significantly from rewriting and clarifying the arguments behind the choices made.

Reviewing the literature, the choice of the dimension of the reduced reference state spanned by the leading EOFs varies from 4 to 14. Many studies used the winter season and the North-Atlantic region, where 4 EOFs already explain more than 50% of the variance (Fabiano et al., 2021, Dawson et al., 2015). Our own sensitivity tests for the extended summer season have shown that 4 EOFs explain about 40% of the variance for the North-Atlantic-Eurasian region and about 34% for the Arctic region. Moreover, we found a stable spatial structure of the regimes obtained with K-Means clustering based on ERA5 MSLP data only when retaining 6 or more EOFs (explained variance larger than 52% and 45%, respectively) for the North-Atlantic-Eurasian region (Figure R4) and for the Arctic region (see figure R5). To ensure a sufficient

amount of explained variance, we decided to consider the 10-dimensional state space consistently for both regions, explaining about 67% of variance for the North-Atlantic-Eurasian region and 61% of variance for the Arctic region.

Please note: We would not include the Figures R4 and R5 in the revised manuscript's appendix but will mention the fractions of explained variance in the revised manuscript.

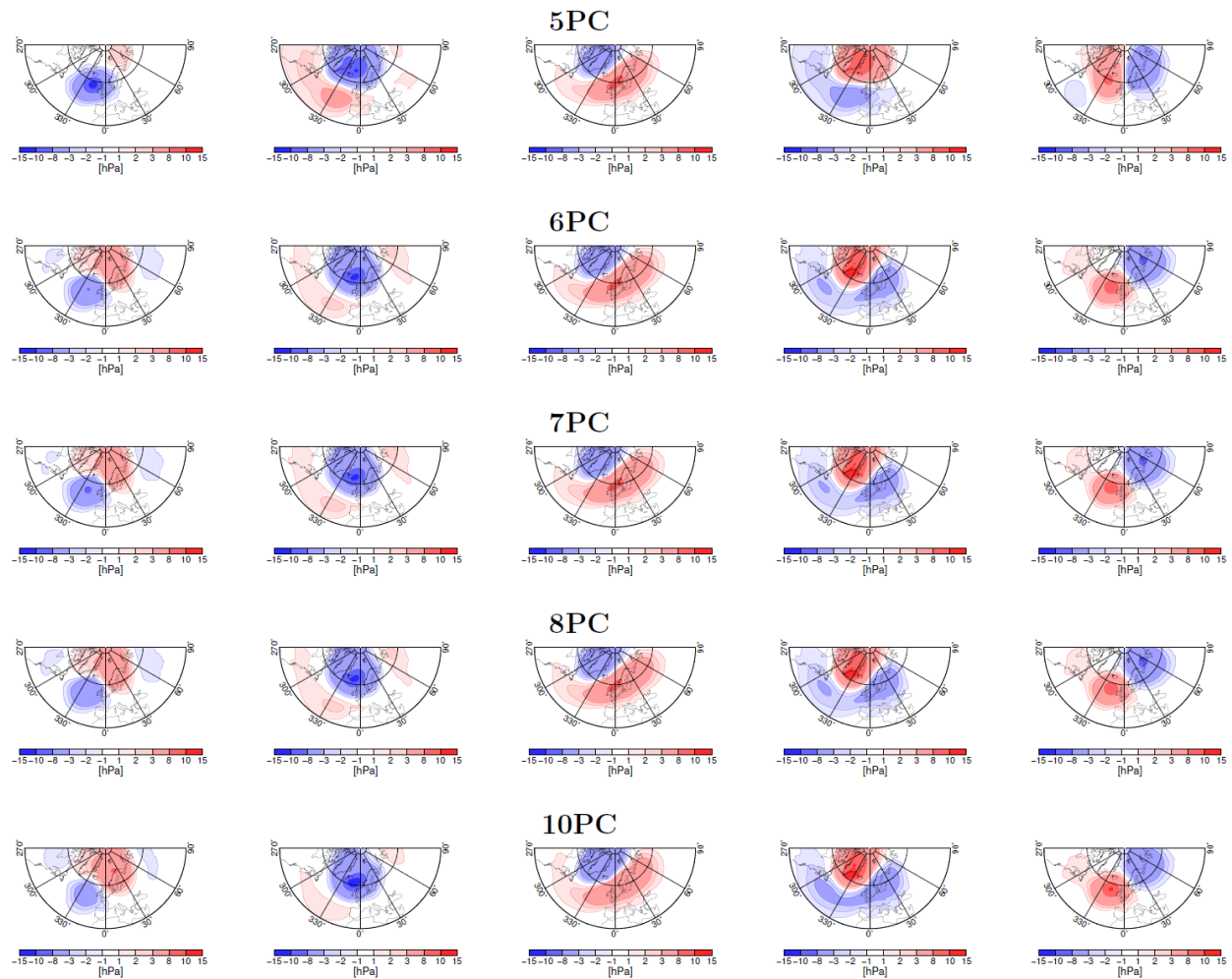



Figure R4: North-Atlantic Eurasian circulation regime patterns in terms of MSLP anomalies for MJJASO, for ERA5 data from 1985-2014 obtained with K-Means clustering in the state space spanned by 5, 6, 7, 8, 10 PCs.

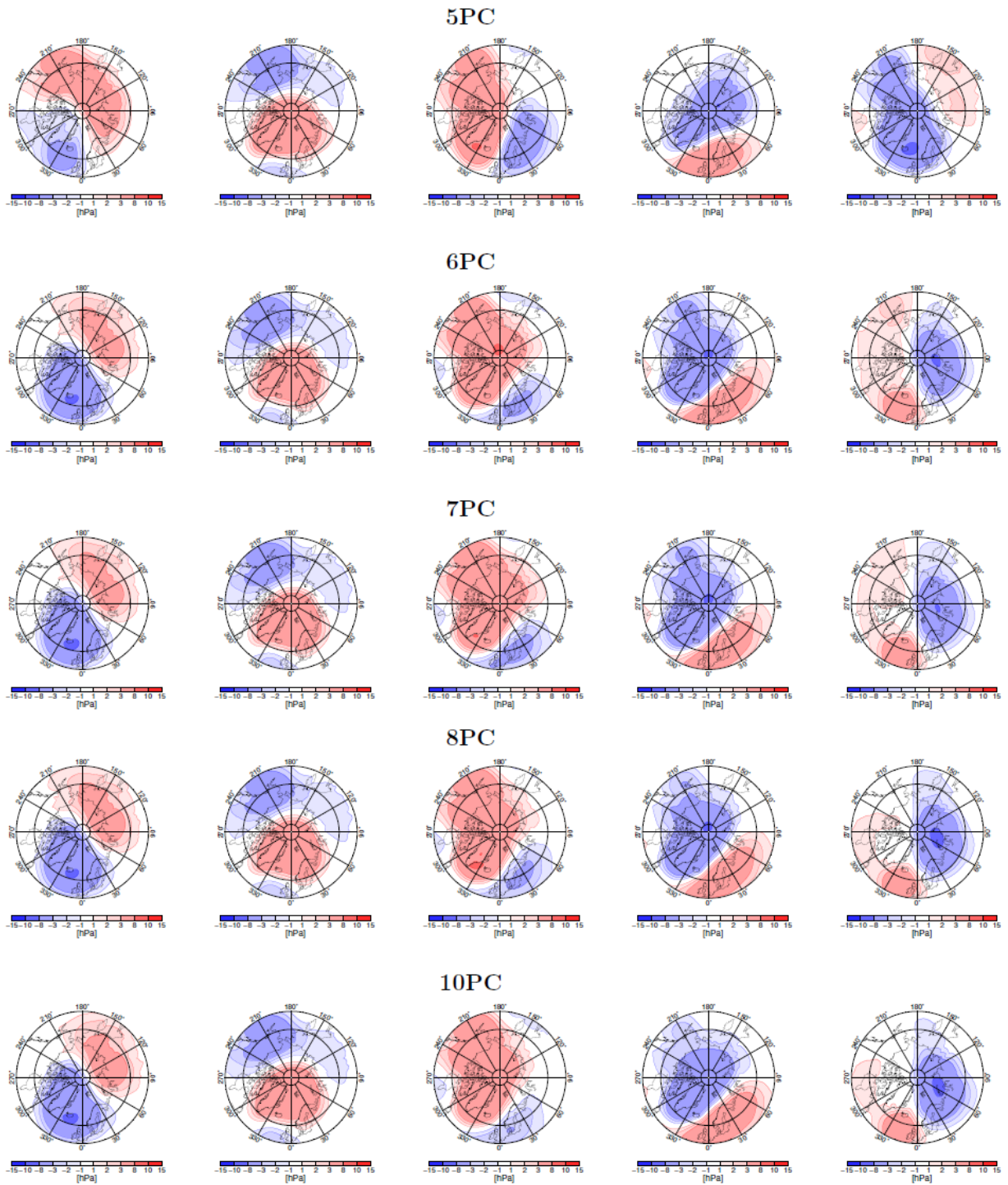



Figure R5: As Fig. R4, but for the Arctic region.

3.The authors study both a projected and simulated approach for k-means, but not for SANDRA. Could you clarify why? A similar question for arguing toward the best number of regimes.

In the SANDRA (SAN) method there is no equivalent of projecting onto the 10-dimensional state space of another dataset. We have now however included a separate calculation of circulation types for the future period from the CMIP models.

We have based the decision for a reasonable number of circulation regimes for the K-Means clustering on the silhouette score and elbow plot metrics described in the appendix. To further justify this choice for both methods, K-Means and SANDRA, here we add an additional metric, following Grams et al. (2017), a reference reviewer 1 suggested in comment 8. As suggested by Grams et al (2017) their metric to evaluate a suitable number of clusters is the Anomaly Correlation Coefficient ("*The optimal number of clusters is seven (Supplementary Fig. 1) based on the criterion that the anomaly correlation coefficient between the clusters is below 0.4*", Grams et al., 2017, Method section). In the following we show the results of the anomaly correlation coefficient calculated for the North Atlantic-Eurasian region in Figure R6 and for the Arctic region in Figure R7 for both methods, KME and SAN, based on ERA5 data 1985-2014. For both methods and both regions, the suggested criteria of Grams et al. (2017) is satisfied for 5 regimes each, which justifies our choices for the number of regimes.

We will include the description of this third metric and the respective discussion in the revised manuscript.

North Atlantic-Eurasian region KME   North Atlantic-Eurasian region SAN

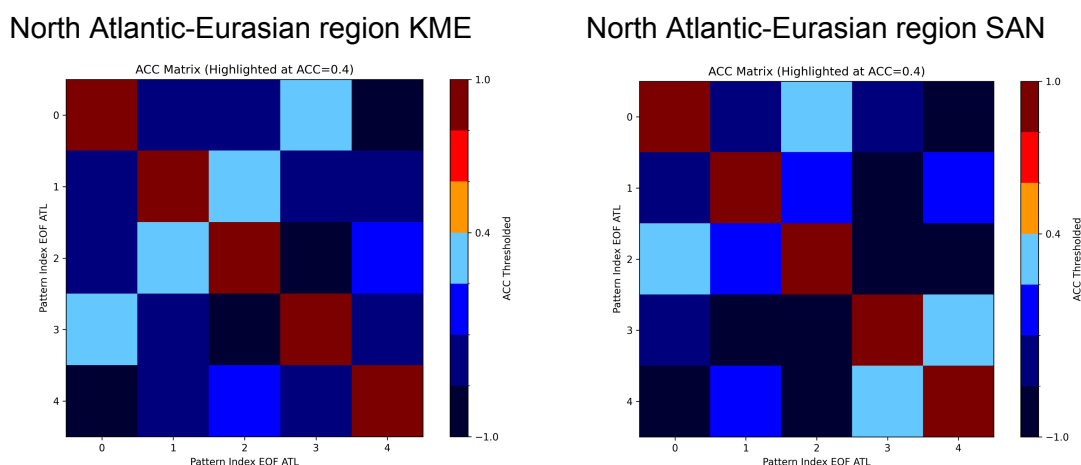
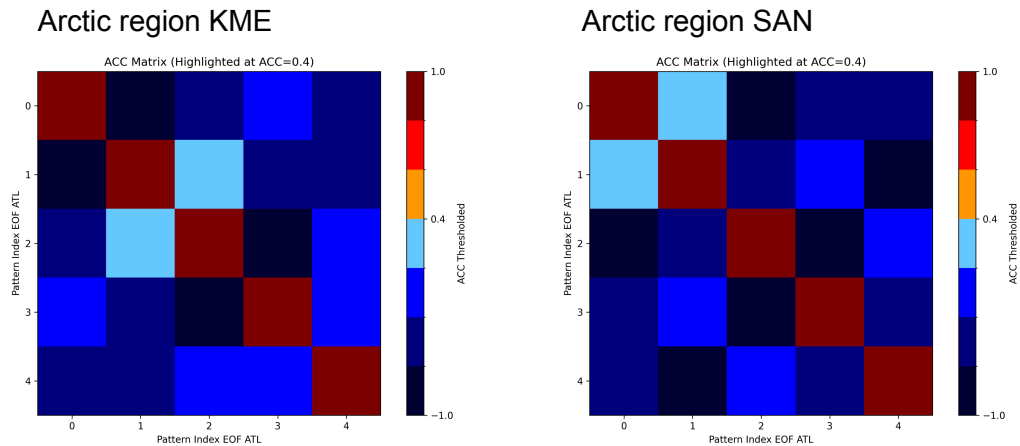



Figure R6: Anomaly correlation coefficient matrix for the North Atlantic-Eurasian regime patterns and both methods, KME on the left side and SAN on the right side. ERA5 data, MJJASO, 1985-2014

<p style="text-align: center; color: green;">Figure R7: As Figure R6, but for the Arctic regimes.</p>

4. In the results the authors argue for just using the projected approach because it is "more accurate", where to me it appears the smaller spread between models is by design. By that I mean that the using each models own regimes by definition will create more variability in the frequency, as now also the regime patterns differ between the models. Can the authors elaborate on this and argue for their choice?

We would argue that while using the models' own regimes would result in a larger variety of the spatial patterns, it would actually lead to a reduction in the frequency bias, as the regimes would be more tailored to each model's circulation states, leading to fewer misclassifications.

However, we see the opposite in the results here, which is why we made the conclusion as we did. The phrasing "more accurate" is perhaps not the best, and will be re-written in the revised manuscript. The more precise phrasing should lean towards the fact that the average and median frequency bias for the projected regimes is smaller compared to the simulated regimes.

We will now try to argue mathematically why this is not by design. If it would be true that this is by design, then the projected approach should lead to smaller frequency biases for **every** model compared to the simulation approach. The frequency bias for every model is given below in Table R1 (for the North-Atlantic-Eurasian regimes obtained with K-Means) and Table R2 (for the Arctic regimes obtained with K-Means). For the ACCESS-ESM1 model (and ACCESS-CM2 for the Arctic) model, the absolute frequency bias for the projected approach is greater compared to the simulated approach. This suggests that the smaller frequency bias in the projected approach compared to the simulation approach is not by design.

Frequency Bias for North-Atlantic-Eurasian regimes, KME

Model	Projected Abs Bias (%)	Simulated Abs Bias (%)
ACCESS-CM2	0.652	1.041
ACCESS-ESM1	1.29	0.483
CESM2-WACCM	0.471	2.446
CNRM-CM6-1	0.442	0.758
CNRM-CM6-1-HR	0.725	1.294
CNRM-ESM2-1	0.239	2.475
GFDL-CM4	0.413	1.41
GFDL-ESM4	0.377	0.903
HadGEM3-GC31-LL	0.449	2.504
INM-CM4-8	0.326	0.823
INM-CM5-0	0.399	1.417
IPSL-CM6A-LR	0.572	2.222
MIROC6	0.406	1.925
MPI-ESM1-2-HR	0.514	2.62
MPI-ESM1-2-LR	0.109	2.381
MRI-ESM2-0	0.203	0.323
NorESM2-MM	1.283	2.345
UKESM1-0-LL	0.681	1.113

Table R1: Frequency bias of projected and simulated approach for North-Atlantic-Eurasian regimes, k-means clustering.

Frequency Bias for Arctic regimes, KME

Model	Projected Abs Bias (%)	Simulated Abs Bias (%)
ACCESS-CM2	1.138	1.028
ACCESS-ESM1	1.144	0.656
CESM2-WACCM	1.384	3.175
CNRM-CM6-1	0.753	2.377
CNRM-CM6-1-HR	0.998	2.514
CNRM-ESM2-1	0.927	3.492
GFDL-CM4	0.528	2.308
GFDL-ESM4	0.897	2.225
HadGEM3-GC31-LL	0.395	2.845
INM-CM4-8	0.567	3.685
INM-CM5-0	0.872	2.157
IPSL-CM6A-LR	0.684	4.607
MIROC6	0.805	5.351
MPI-ESM1-2-HR	0.521	3.836
MPI-ESM1-2-LR	0.815	5.805
MRI-ESM2-0	0.613	0.67
NorESM2-MM	1.496	2.308
UKESM1-0-LL	1.023	1.909

Table R2: Frequency bias of projected and simulated approach for Arctic regimes, k-means clustering.

5. The authors repeatedly mention the patterns being similar between the two methods considered (KME and SAN), where looking at the figures I do not think the differences are non-significant for all regimes. This is also indicated by the different response to climate change of some regimes. I do not have an intuition about how to interpret the SANDRA regimes with respect to those of k-means. It would be helpful if the authors could elaborate on the difference between the methods and its impact on the regime interpretation.

Our sincere apologies! We have discovered an error in our SANDRA (SAN) setup, and it turned out that area weighting was turned off. This gave an unproportionally large weight to the grid cells near the pole, which of course affects the patterns. We have now run SAN again with area weight turned on (by the cosine of the latitude, just as for the KME method) and the new results indeed look a lot more similar between KME and SAN.

Regarding interpretation, SAN, like any other optimization method, including KME, uses statistical measures to obtain a set of regimes that are the optimal representation of the dataset in terms of metrics such as minimum inter-cluster distance. There is no inherent feature in the methods that would suggest they favour certain patterns over others.

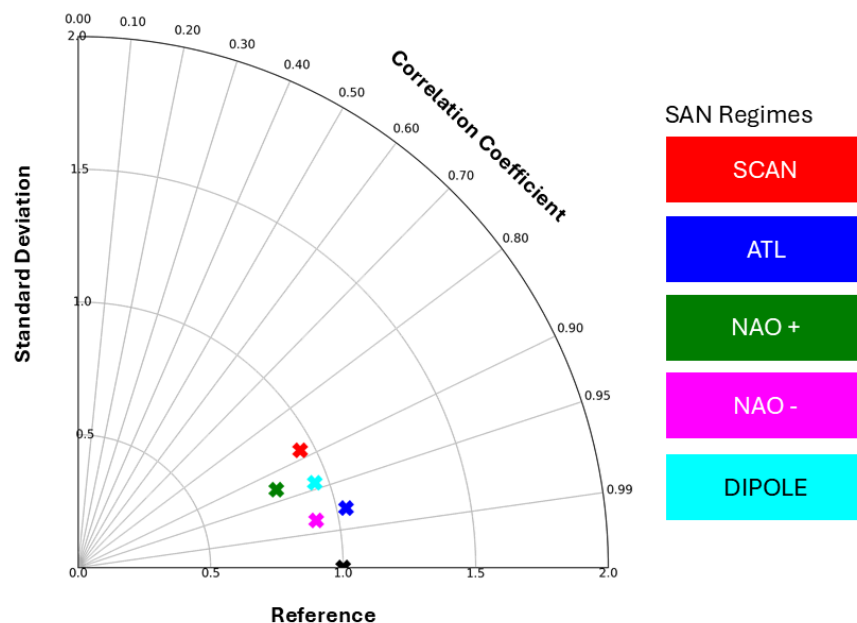



Figure R8: Taylor diagram analysis of circulation regimes computed by KME and SAN algorithms in the historical period for the North Atlantic-Eurasian region in extended boreal summer season from May to October. The reference circulation regimes computed from ERA5 reanalysis data with KME is marked as black cross, the colored crosses represent the SAN regimes.

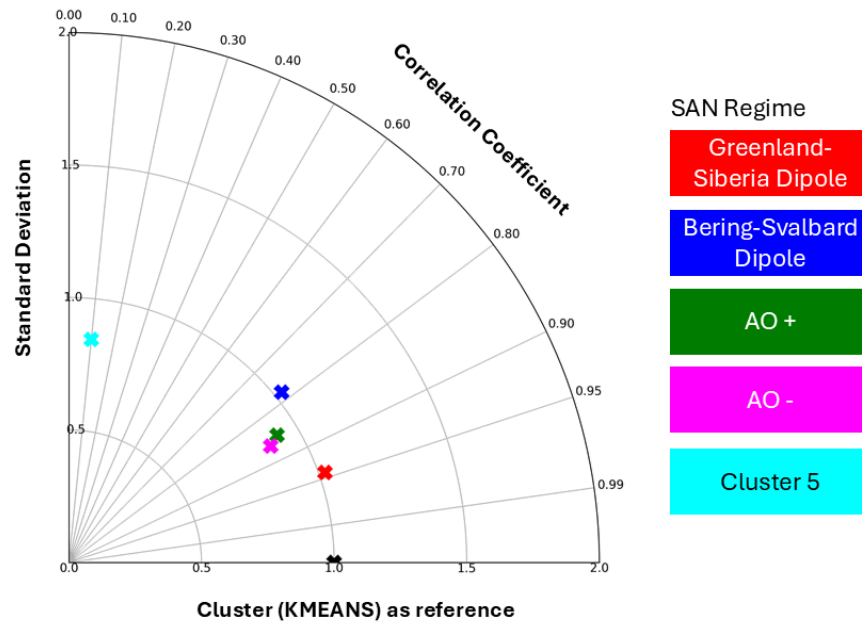



Figure R9: Taylor diagram analysis of circulation regimes computed by KME and SAN algorithms in the historical period for the Arctic region in extended boreal summer season from May to October. The reference circulation regimes computed from ERA5 reanalysis data with KME is marked as black cross, the colored crosses represent the SAN regimes.

A common result when applying different circulation type classification methods is that while some patterns appear very similar between different methods, but the small differences will lead to differences in which days will be assigned to which pattern, and in the end this often results in one of the patterns becoming more different. As an example, for the Arctic patterns (Fig. R15 and R16), we note that the low pressure in the "Bering-Svalbard dipole" cluster is located over Alaska in KME, while it is more towards Kamtchatka in SAN. This is compensated for in SAN by grouping together days with low pressure over Alaska in cluster 5. We will discuss this in the revised manuscript.

Including the updated results from SAN, Figure R8 and Figure R9 evaluate the spatial similarity between KME and SAN results in a Taylor diagram. For the North-Atlantic-Eurasian regimes (Figure R8), all pattern correlation coefficients are high (> 0.88 for all patterns) with similar spatial standard deviations, underlining that both methods revealed similar regimes. For the Arctic, 4 out of 5 regimes are detected similarly (pattern correlation >~0.8, similar spatial standard deviation). Cluster 5 is very different, as explained in the previous paragraph.

Accordingly, we updated the figures showing the changes in occurrence frequency under climate change (Figures 6 and 9 in the old manuscript, updates in Figures R17 and R18.)
With the updated SAN results, there is a strong agreement between changes in frequency for the North Atlantic domain, with all 5 regimes having the same sign of the ensemble median of frequency changes. For the Arctic however, only two of the regimes show robust results when comparing the methods (decrease of AO+ and increase of AO-).
While it is natural that a difference in the circulation patterns also lead to differences in frequency, in the updated manuscript we will also add a reflection on the robustness, e.g.

"The fact that while patterns 1-4 for the two methods in the Arctic appear similar, even small differences in the patterns lead to notable differences in the change in frequency [Fig. R18]. Only the AO+ and AO- have the same ensemble median change in frequency, leading us to conclude that changes in the occurrence frequency in these regimes are more robust than changes in the others."

6.Figure 4. I do not know what is shown in this figure. Which regimes are compared to the reference? I would say each model has their own, whereas the authors refer to the common EOFs of CMIP6 models. I could not find anything in the methods section discussing this. I think it would be valuable to show the spread between models in a figure like this, also finding which models have a regime representation that is closer to reality.

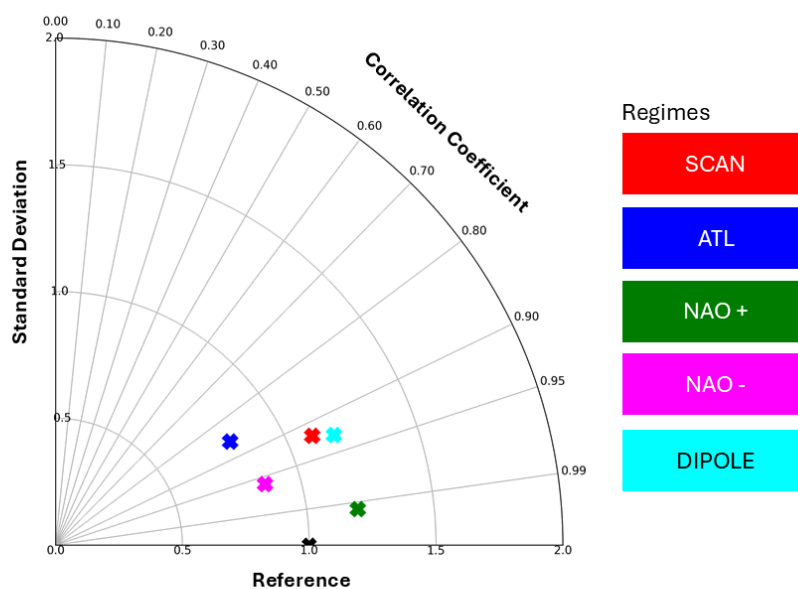



Figure R10 (Figure 4 in original manuscript): Taylor diagram comparing the future common simulated regimes for the CMIP6 ensemble for SSP5-8.5 and the period 2070–2099 to the corresponding reference regimes from ERA5 reanalysis in the historical period (1985–2014), for the North Atlantic-Eurasian region. Spatial regime patterns determined with K-Means clustering.

Using Figure R10 (Fig. 4 in the old manuscript) we evaluate the future changes in the spatial structure of the circulation regimes by comparing the future common simulated regimes for the CMIP6 ensemble for SSP5-8.5 and the period 2070–2099 with the ERA5 reference regimes for the historical period 1985-2014. for the North Atlantic-Eurasian region. The respective Taylor plot for the Arctic regimes is shown in Fig. R20 (and will be included in the revised manuscript).

Both Taylor plots, Fig. R10, R20, underline the similar spatial structure of the common simulated future CMIP6 regimes to that of the ERA5 regimes for the historical period 1985-2014, with all pattern correlation coefficients higher than 0.8 food the North Atlantic-Eurasian and higher than about 0.8 for the Arctic regimes. In our view this justifies the application of the "projected

approach" (see our terminology described in Fig. 1 of the manuscript) for the calculation of future changes in regime occurrence frequency (and persistence).

Note: The description of the common approach (compare answer to Rev.1, comment 1), and of the Figure caption to Figure 4 will be improved in the revised manuscript.

7.The authors only look at changes in the regime occurrence. It would be valuable to also study changes in the persistence of the regimes, to get a more dynamical understanding of the effects of climate change, e.g. is a regime becoming more frequent because it becomes more persistent? I suggest adding results on persistence.

We thank reviewer 1 for this comment and we will add results on persistence in the revised manuscript. The new results for the North Atlantic-Eurasian regimes are shown in Figure R11, those for the Arctic regimes in Figure R12.

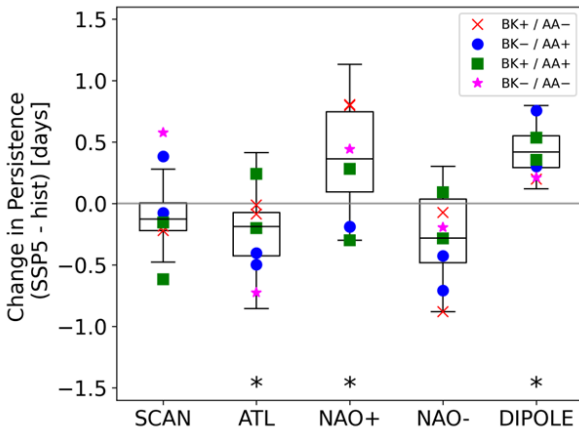
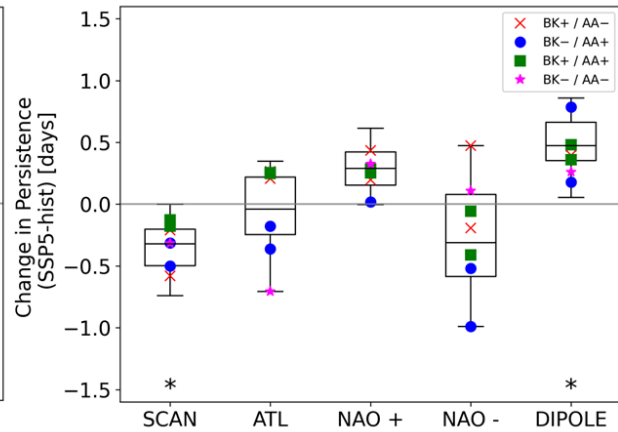
Indeed, most of the changes in the frequency of occurrence of the North Atlantic-Eurasian regimes (refer Figure R17) coincide with the changes in persistence. For KME and SAN, the projected changes under SSP5-8.5 scenario are similar for both metrics (persistence and frequency). But the changes are not significant for the projected persistence for Scan and NAO-regimes regarding KME and not significant for NAO+ and NAO- regimes regarding SAN method.

KME   SAN

Figure R11: Changes in the persistence for the North Atlantic-Eurasian region under global warming (SSP5-8.5 scenario) compared to the historical period, in the extended boreal summer season May to October. The boxes denote the first and third quartiles, the center black line indicates the ensemble median and the top and bottom whiskers represent the 10th and 90th percentiles. Stars indicate significant changes compared to the historical data at the 95% confidence level calculated with Welch's t-test. Markers state the models associated with storylines.

Analysing the Arctic regimes, the projected changes in persistence are not as aligned with the frequency in occurrence, refer to Figure R18 but still coincide with the results in the projected frequency changes under SSP5-8.5 scenario for some regimes. Regarding the KME method, both metrics show similar significant changes for AO+, AO- and Cluster 5. Comparing the shifts for the Greenland-Siberia Dipole and Bering-Svalbard Dipole, persistence and frequency show opposing results, those changes are not significant for persistence. For the SAN algorithm, the projected significant positive change in the frequency of occurrence for AO- pattern is confirmed by the projected persistence change. Other projected persistence changes are not significant but show a similar tendency compared to the frequency changes except for Cluster 5. Here the frequency of occurrence is simulated to decrease significantly under the SSP5-8.5 scenario but the persistence is increased.

KME

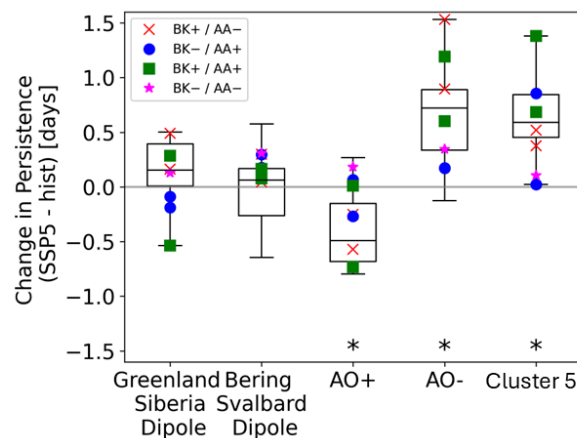
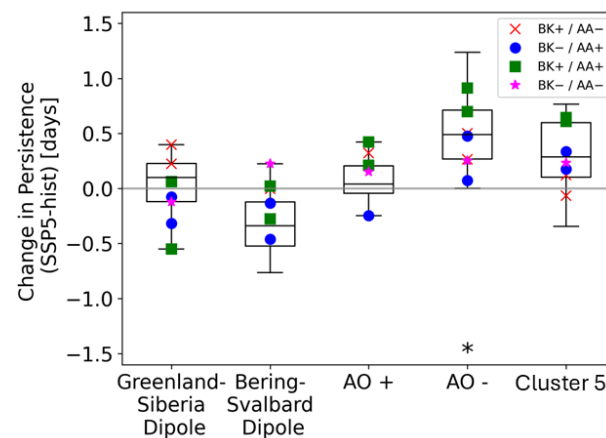
SAN

(a) KME   (b) SAN

Figure R12: Changes in the persistence for the Arctic region under global warming (SSP5-8.5 scenario) compared to the historical period, in the extended boreal summer season May to October. The boxes denote the first and third quartiles, the center black line indicates the ensemble median and the top and bottom whiskers represent the 10th and 90th percentiles. Stars indicate significant changes compared to the historical data at the 95% confidence level calculated with Welch's t-test. Markers state the models associated with storylines.

8.Many of the studies references for comparison study different domains, e.g. Europe in Boé (2019). It would be good to acknowledge that in the text, as it can impact the results. The same for studies used in comparison that study the winter season. Furthermore, I suggest adding some references on the different number of regimes:
https://www.nature.com/articles/nclimate3338 and
https://rmets.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/qj.3818.

Both studies will be cited in the revised manuscript. Similar to Grams et al. (2017), we introduced the anomaly correlation coefficient to justify the chosen cluster number of 5 for both regions and both methods (see above our answer to comment 3). The second paper will also be included when discussing the general debate about an optimal cluster number.

In the revised manuscript we will be more accurate regarding seasonality and regionality.
In particular the text L193-209 will be changed to:

"The reference circulation regimes for the ERA5 data 1985-2014 obtained with the K-Means clustering (KME, Fig. 2 upper row) and the SANDRA (SAN, Fig. 2 lower row) algorithm are summarised here:

– The Scandinavian Blocking regime (Scan), shown in Fig. 2a, indicates a positive pressure anomaly centered above Scandinavia and a low pressure anomaly, centered over the North Atlantic Ocean westwards of the British Isles (refer to KME, Fig. 2a upper row). The regime pattern obtained by the SAN algorithm displays a slightly weaker positive pressure anomaly over Scandinavia and a slightly stronger low pressure anomaly over the North Atlantic (Fig. 2a, lower row).

– The Atlantic Trough regime (Atl) is characterized by a strong negative pressure anomaly centered between Iceland and the British Isles (Fig. 2b).

– Fig. 2c displays the North Atlantic Oscillation, which is poleward shifted in summer, in its positive phase (NAO+). Here an elongated positive pressure anomaly extends from the Ural to Newfoundland accompanied by a negative pressure anomaly above Greenland.

– Fig. 2d displays the North Atlantic Oscillation in its negative phase (NAO−), here the positive and negative pressure anomalies are swapped respective to NAO+.

– The Dipole Atlantic Blocking regime (Dipole, Fig. 2e) displays a positive pressure anomaly centered between Iceland and the British Isles. The accompanying negative pressure anomaly is located over the Ural region and Barents Sea.

 Riebold (2023) analysed circulation regimes over the same region for the summer season (JJA: June to August) and found similar MSLP anomaly patterns for each regime. Boe et al. (2009) and Cattiaux et al. (2013), on the other hand, analysed summer circulation regimes for JJA over a smaller North Atlantic region (20°–80°N, 90°W–30°E), using MSLP (Boe et al., 2009) and Z500 (Cattiaux et al., 2013) data, respectively. In both studies, four circulation regimes were detected. These four regimes comprise the negative and positive phases of the NAO (called NAO- and Blocking, BL), an Atlantic low (AL) and an Atlantic ridge (AR) regime, which bear many similarities with the NAO-, NAO+, Atl and Scan regimes detected in our study. "

In addition, we checked our references to studies analysing winter regimes and found that we had properly referenced them as such.

9.I do not see the added value of discussing storylines for this paper. Since each storyline only contains two models, it is near-impossible to draw any robust conclusions. Can the authors clarify why they chose to also include the storylines in this work?
The main motivation behind adding the storyline perspective is to put the work in context with other ongoing work in the EU PolarRES project, which will also see more studies being

published in the coming year. Regarding the circulation response to storylines, even with the small sample, there at least seems to be a tendency of the models of each storyline to be closer to each other in the same half of the box plots.

As shown in Levine et al. (2024, fig. 4), the CMIP6 multi-model mean (MMM) shows a poleward shifted N-Atlantic jet, but the strength of this poleward shift depends on the storyline, with a weaker poleward shift in storyline A and B and slightly stronger poleward shift in storyline C and D (compared to the MMM). Since the circulation regimes are related to specific patterns of jet anomalies, one might hypothesize that this is related to different future changes in the occurrence frequencies of in particular those regimes which are related to a northward shifted jet (NAO+, Dipole). We are aware that testing this hypothesis is limited because there are only a few models available which represent each storyline. (These paragraphs are also used as answer to reviewer 2, comment 2.)

We now consider models for all 4 storylines and will discuss this in the revised manuscript.

10. The conclusion and outlook section could benefit a lot from placing the work in the context of the wider literature. Are the finding in line with other studies, with what we expect linked to jet stream shifts, … As is, the outlook part is very brief and lacks context.

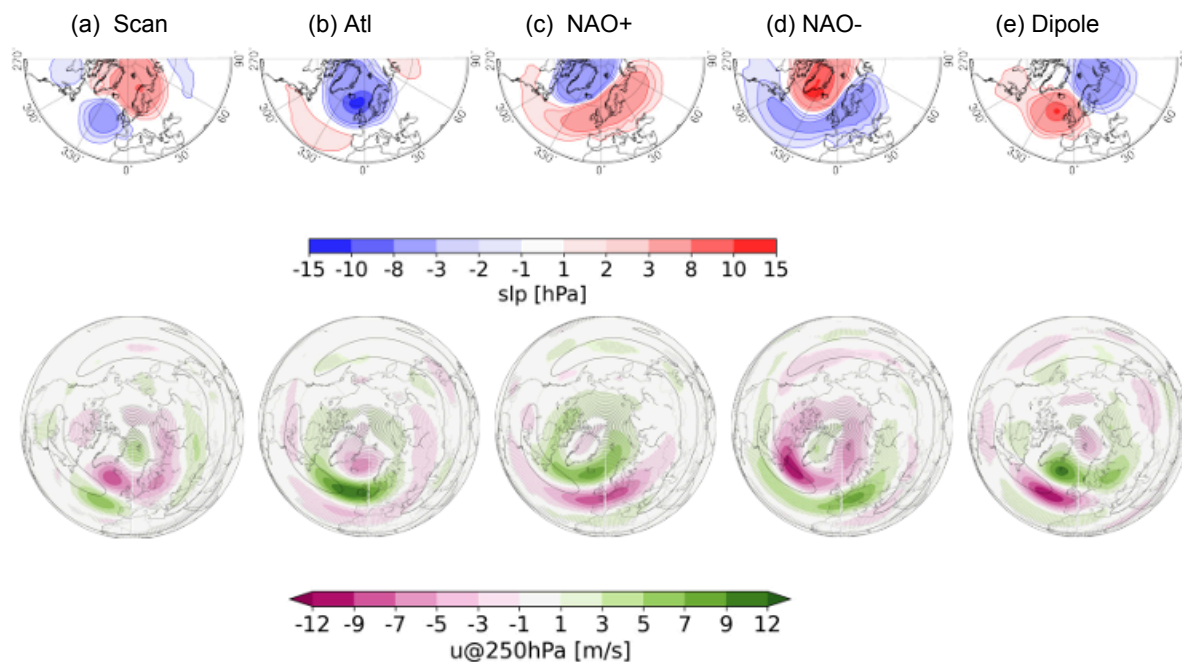



Figure R13: MSLP anomaly fields for North-Atlantic-Eurasian reference regimes (from ERA5) computed with KME method in upper row. Zonal wind composites @250hPa level in lower row. Climatology is plotted by black contour lines.

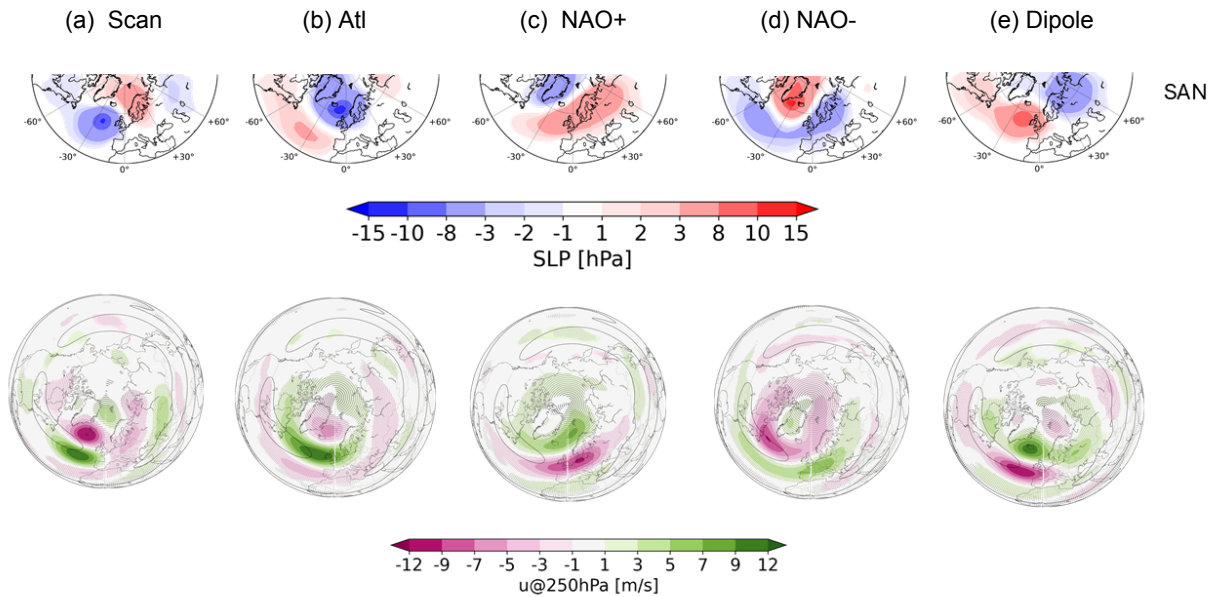



Figure R14: As R13, but for regimes obtained with SAN method.

We have done some analysis on the changes in the wind field, see below. We will add the following discussion as a new subsection to the results section which includes relations to existing literature on the subject.

To relate our results of regime changes changes in tropospheric jets, Figures R13 and R14 show composites of zonal wind anomalies at 250hPa (u@250hPa) for the North-Atlantic-Eurasian regimes determined with KME (Figure R13) and SAN (Figure R14). Due to the similar structure of the regimes obtained with both methods, the composites of u@250hPa display similar patterns, too. The Scan and NAO- regimes are associated with a southward shifted jet above the Atlantic. NAO+ and Dipole are linked to a poleward shifted jetstream. The NAO+ regime is also connected to a slightly tilted jet exit. The Atl regime is associated with a strengthening of the jet exit. The predicted northward shift of the jet in the North Atlantic detected by Woollings & Blackburn (2012) (CMIP3) and Barnes & Polvan (2013) (CMIP5), Levine et al. (2024) (CMIP6) and Harvey et al. (2020) (CMIP3,5,6) is consistent with the increase in occurrence of the NAO+ and Dipole regimes under SSP5-8.5 scenario. In accordance, both regimes that are linked to a southward shifted jet (Scan and NAO-) are simulated to occur significantly less frequently in the future. The future increasing tilt of the jet in the Atlantic region (Woollings & Blackburn, 2012) is in agreement with the detected increase in the frequency of occurrence of the NAO+ regime in our study.

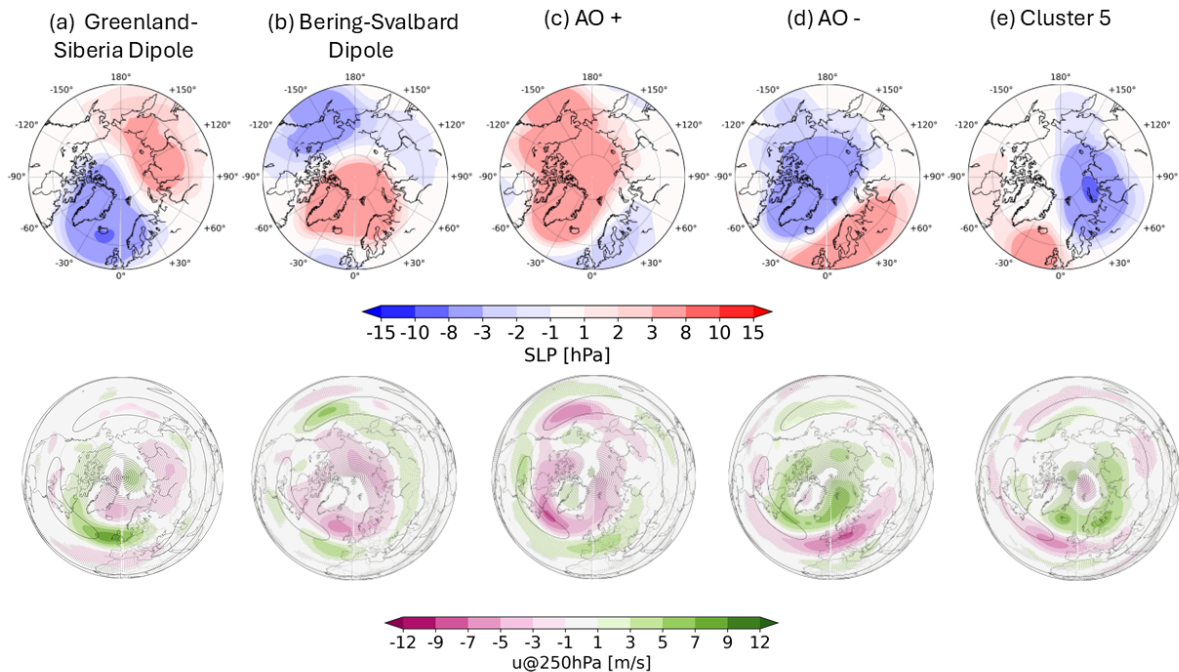



Figure R15: As R13, but for the Arctic regimes.

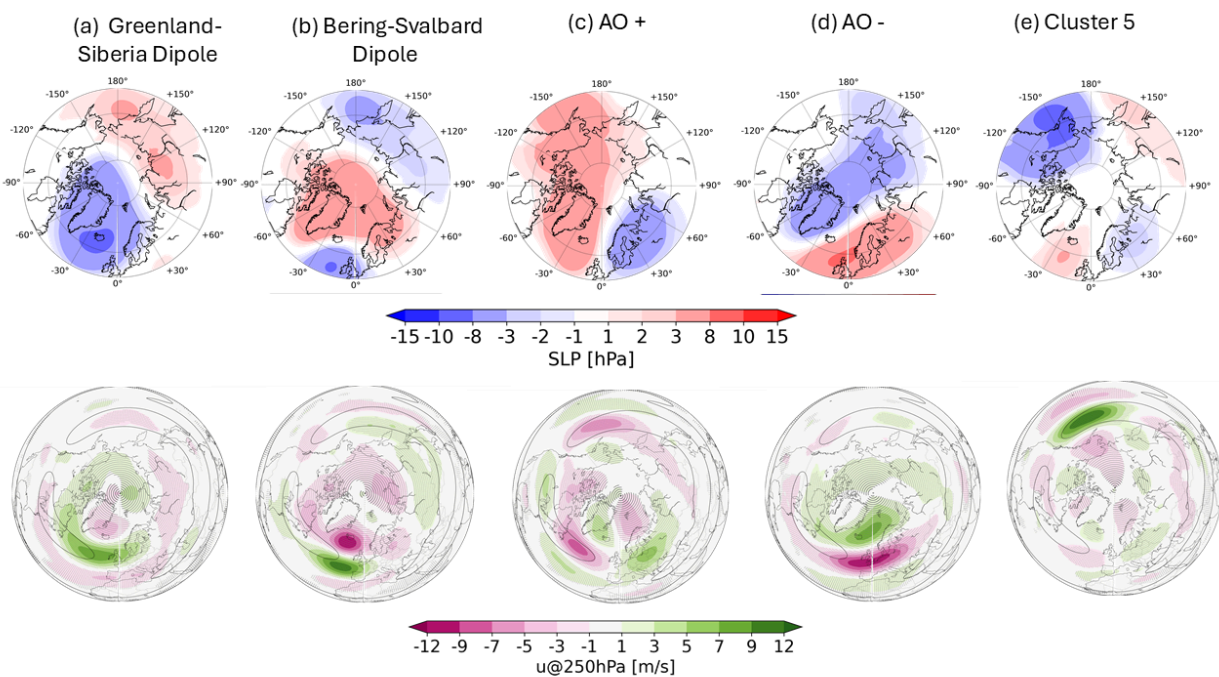



Figure R16: As R15, but for regimes obtained with SAN method.

The respective u@250hPa composites for the Arctic regimes are shown in Figure R15 for the KME method and in Figure R16 for the SAN method. These composites (Figure are again very similar for those 4 regimes with similar pattern structures in KME and SAN (Figures R15 (a)-(d) and R16(a) -(d)), but very different for cluster 5, compare Figures R15 (e) and R16(e).The u@250hPa composites show a strengthening of the Atlantic jet exit for Greenland-Siberia Dipole regime. The Bering Svalbard Dipole regime is associated with a southward shifted Atlantic jet, similar to the Scan pattern. The AO+ regime is associated with a weakening of the Atlantic and Pacific jet exit, with a slight southward shift for the North Atlantic jet, similar to NAO-. In contrast, the AO- regime is associated with a northward shifted North Atlantic jet, similar to NAO+. The most robust change in the occurrence frequency we detected is the increase in occurrence of the AO- regime, i.e. a poleward shift of the North Atlantic jet, which is in accordance with the above cited studies. In accordance, the AO+ regime and the Bering Svalbard Dipole (for KME only) which are linked to a southward shifted North Atlantic jet, are simulated to occur significantly less frequently in the future.

Comments from the PDF:

L8: "Specify which regimes".

> We rewrite "Despite slight differences between the SAN and KME methods in identifying spatial regime structures, the ~~fundamental spatial configuration of these regimes~~ patterns remain~~s~~ largely unchanged under future climate scenarios."

L22: "or" -> "and/or"

L24: "not change the spatial structure of the regime patterns" comment: "Needs a reference. It's the ones you give after, as there's not much, but which specifically do you refer to?"

> We are adding Palmer (1993, 1999) already here.

L25: "strong"

> See response to reviewer 2, comment 4.

L37 "storyline approaches", comment "describe in one sentence"

> Moving from L58.

L51: "Is this CMIP?", comment "Is this CMIP?"

> We add ", CMIP3"

L58: "They provide" -> "Mention where you first mention storylines"

> Thank you for the suggestion. We are moving this to L37.

L66: "projected future changes", comment "Be more specific. Changes in what?"

> We are adding "with focus on frequency of occurrence and persistence".

L68: "the results", comment "Which results?"

    We are adding "frequency of occurrence as well as persistence".

L73: "Only SSP5-8.5?"

    We clarify and shorten by rewriting "In examining the simulation of future time periods, the SSP5-8.5 emission scenario is of particular interest. As the highest emission scenario, it is characterised by a large increase in Greenhouse Gas (GHG) emissions, impinging a strong external forcing upon the atmosphere. Thus, a response of characteristics of circulation regimes is expected." to "For the future period we use the highest emission scenario SSP5-8.5 as it is expected to give the most statistically robust response to climate change."

L79: "Why SLP and not Z500?"

    Atmospheric circulation regimes are generally computed by performing clustering algorithms on a circulation variable such as the geopotential height at 700 or 500 hPa or mean sea level pressure MSLP). We agree that the majority of studies on atmospheric circulation regimes is based on analyses of the the geopotential height field at 500hPa, but there are also many studies analyzing geopotential height field at 700hPa (e.g. Michelangeli et al., 1995, Vautard 1990) or mean sea-level pressure (MSLP) fields (e.g. Philipp et al., 2007, Boe et al., 2009, Crasemann et al., 2017, Skific et al., 2009) .

    Here, we chose to use MSLP instead of geopotential height as the future trend is much smaller in the MSLP fields.

L80: "(30◦ N-90◦ N, 90◦ E-90◦ W)", comment "This is larger than in most other studies on regimes in the Euro-Atlantic sector".

    We will reformulate this sentence to "Given that the circulation regimes in this area have been studied before (Crasemann et al. 2017) during winter, we now want to study this area in the summer season."

L85: "Is 30 years sufficient data?"

    This is commonly used in multiple climate studies.

L85: "Sentence too long and difficult to follow."

    We split the sentence after 2099.

L93: "To the seasonal cycle?"

    Yes. We clarify by reformulating as "The mean seasonal cycle is obtained by computing the daily climatology and then applying a 21-day moving average to this."

L94: "influenced by GHG forcing", comment "I understand that, but that does not fully answer the question as to why you use the historical one also for the future. It'd mean you consider the changes in the seasonal cycle as a part of the changes in regime dynamics."

    We clarify by changing

"This difference is taken into account by subtracting the historical seasonal cycle from both the historical and future time periods of climate model simulations as presented in Fabiano et al. (2021)" to "To keep this dynamically induced change in the future seasonal cycle, we subtract the same historical seasonal cycle from both the historical and the future periods, as also done in Fabiano et al. (2021)."

L97: "interpolated onto the identical 1.125x1.125 grid", comment: "Why? As this is not necessarily a requirement for the PC/EOF computation, that will work fine for any grid. Not sure about SANDRA. Do you think it could introduce an additional bias?"

The CMIP6 models have different resolutions, so for the common EOF part of the methodology, they need to be interpolated to a common grid. The choice of 1.125x1.125 was rather arbitrary and taken from previous studies, but as also shown, the results are not sensitive to using the 2.5x2.5 degree grid.

L104: "the ten-dimensional reference state space", comment "What do you mean by this?"

We rephrase this sentence as
"The first ten EOFs (computed from our ERA5 reference dataset) constitute the basis vectors for our ten-dimensional reference state space."

L111: "anomalies were projected onto the ten-dimensional reference state space", comment "Why? Provide some reasoning"

Refer to reviewer 1, comment 1 and reviewer 2, comment 1.

L115: "projected approach", comment "So in summary, you assign the model data to the ERA5 regimes, in a nuanced way."

Yes, to clarify we add "In other words, we assign the data from the CMIP models to the regimes from ERA5 by using the pseudo-PCs." in L116.

L139: "How do you interpret the patterns that come out of SANDRA?"

See our answer to comment 5.

L148: "projected approach for SANDRA", comment "Why?"

See our answer to comment 3.

L153: "coarser resolution", comment "Ok, but then you might as well also use this for the EOF/PC computation, as there I wouldn't expect much difference either,."

The choice of 1.125x1.125 was rather arbitrary and taken from previous studies, but as also shown, the results are not sensitive to using the 2.5x2.5 degree grid. In our SANDRA setup we had previously mostly used 2.5x2.5, so we also ran tests with both resolutions to verify that there was no significant difference in the results.

L156: comment: "Add Falkena et al. (2020)"

We will add this reference.

L158: "The same for both KME and SAN?"

  See comment 3.

L162: "Further Methods", comment "Be specific in the heading what this is for."

  We will change this heading to "Statistical Evaluation Methods".

L177: "Add references".

  We will add the storyline references mentioned in the introduction also here.

L183: "Why determined by opposite signals, and not also ++ or --?"

  We are adding the other two storylines too, see reviewer 2, comment 2.

L186: For section 4.1, "Subsections in here would help the reader."

  We agree. We are adding "Spatial regime patterns for the present day" and "Future changes".

Fig. 2, comment: "Why here all capitals and not for Arctic region?"

  We will correct this in the revised version.

L195: "What do you mean by centre of action?"

  We will replace this with "pressure anomaly centres".

L202: "Change brackets."

  We will correct this in the revised manuscript.

L230 and comment on Fig. 2

  See our response to comment 5.

L241: "common simulated circulation regimes", comment "What are these?", and comments to Fig. 4.

  We will try to describe this better in the revised methods section. The common approach is to compute one set of EOFs for all CMIP models taken together as one dataset.

L247: "How would this be different if you use the model PC's instead of pseudo-PC's?"

  See our answer to comment 1.

L251: "I find this too strong a statement because of the use of pseudo-PC's in the methods."

  See our answer to comment 1.

L255: "more accurately", comment "What do you mean by more accurately? It's a different approach, and I would expect this difference given that for the simulated approach the regimes differ between models whereas for the projected they don't..."

  See our answer to comment 4.

L280: "I struggle to see the value of the storylines if you only have two models for each."
See our answer to comment 9.

L286: "diverging future changes", comment "Because the patterns are different would be my hypothesis, elaborate a bit."
We agree, and will elaborate more on this discussion in the revised version.

L289: "small differences", comment "Not always that small."
Results are updated and the differences are smaller, see our answer to comment 5.

L295: "tendency of models associated with BK+/PA− storyline to occur even more often" comment "One is on the median, so you're basing this on one model, which I think is a too strong statement."
We agree, and wording will be updated in light of the new results.

Fig. 6, comment "Difference for DIPOLE, also larger spread for SAN"
Figures are updated (see Fig. R17, R18) and text will be as well.

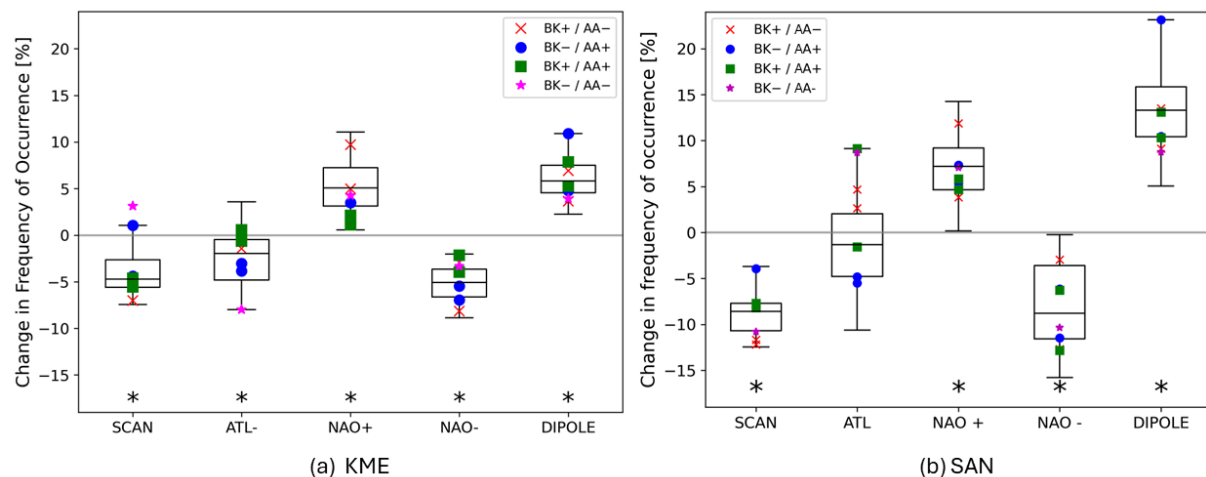



Figure R17: (Updated Fig. 6 original manuscript): Changes in the frequency of occurrence for the North Atlantic-Eurasian regimes under global warming compared to the historical period, SSP5-8.5 scenario in the extended summer season May to October. The boxes denote the first and third quartiles, the center black line indicates the ensemble median and the top and bottom whiskers represent the 10th and 90th percentiles. The star indicated significant changes compared to the historical data at the 95% confidence level calculated with Welch's t-test.The markers represent models attributed to the respective storylines.

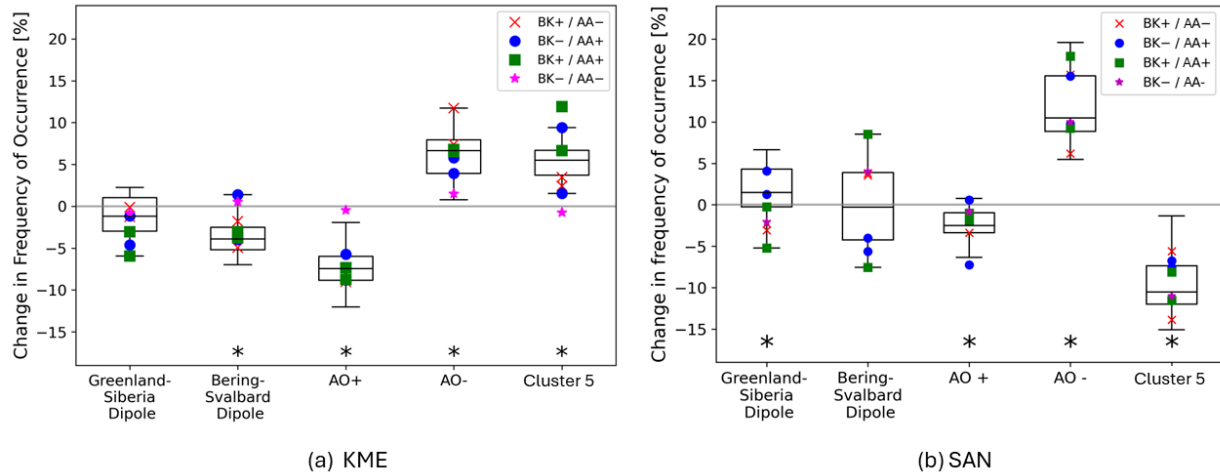



Figure R18: (Updated Fig. 9 original manuscript) As Figure R17 for the Arctic regimes.

Fig. 7, comment "Also, AO+/- are very different in it's extent. And E-W Dipole is substantially different as well",
L318, "Again I find this too strong a conclusion, even more so than in the other domain.",
and L365: "similar in their spatial structure", comment "Questionable".

> We agree. The patterns have changed and the discussion will be adapted accordingly. See comment 5.

L320: "No analysis of pattern change in the future?" and L377: "You haven't shown this for the arctic regimes."

> Thank you for this comment. We had done this analysis as well, and we will include the respective Taylor plot Figure R20 and its description in the revised manuscript. Figure R20 reveals the similar spatial structure of the common simulated future CMIP6 regimes to that of the ERA5 regimes for the historical period 1985-2014, with all pattern correlation coefficients higher than about 0.8 for the Arctic regimes. In addition, here we show the respective Arctic regime pattern for the ERA5 regimes for the historical period 1985-2014 (Fig. R19, upper row) and the common simulated CMIP6 regimes for the future. Also the visual inspection reveals the similarity of the patterns, with some changes in the strength of the MSLP anomaly centers (as for AO+ and Greenland-Siberia Dipole) or in the position of the MSLP anomaly centers (as for Cluster 5).

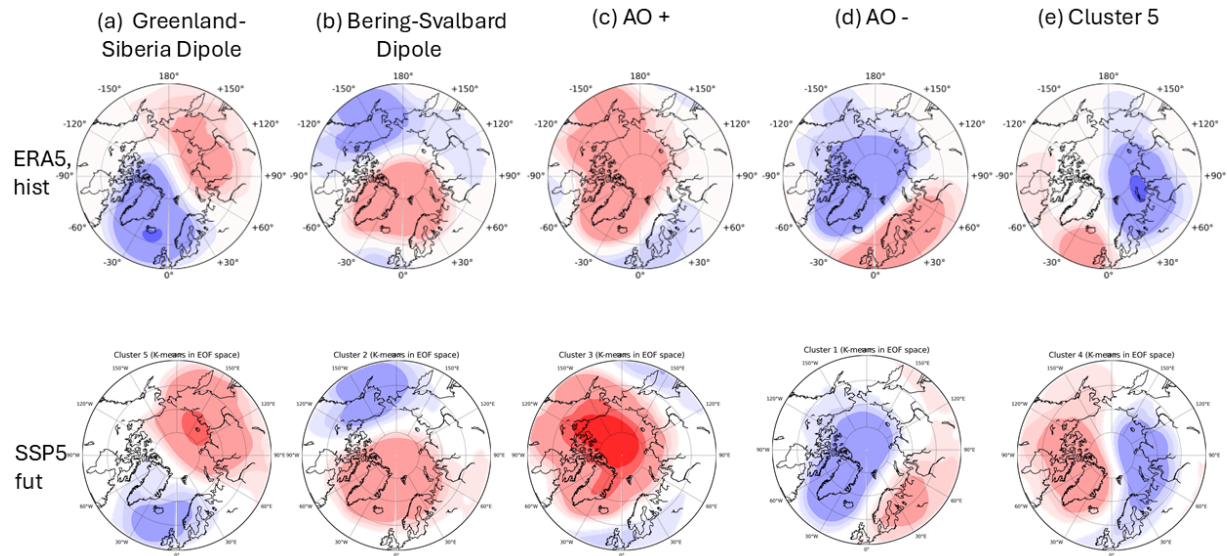



Figure R19: MSLP patterns computed by KME method for ERA5, historical time period in the upper row and common circulation regimes derived from CMIP6 models in the future period (SSP5-8.5, 2070–2099) in the lower row, Arctic region.

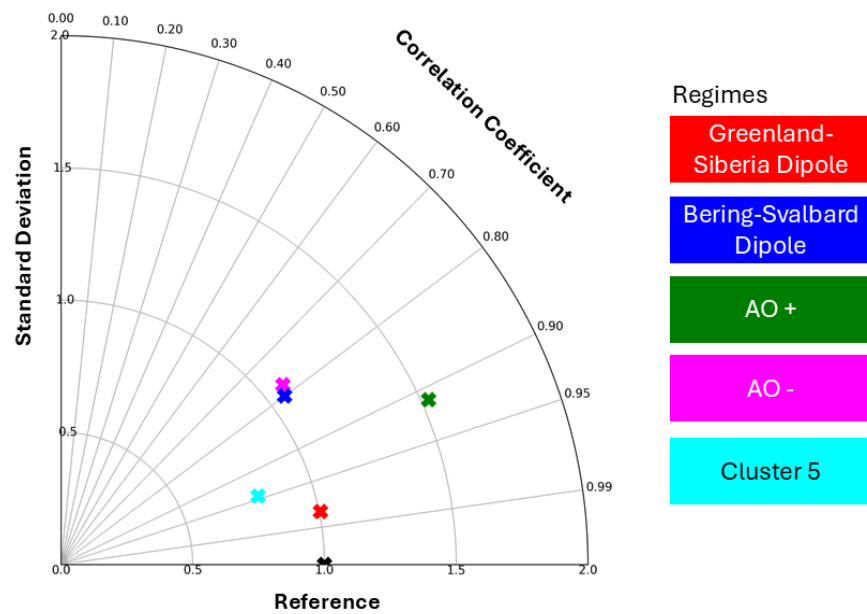



Figure R20: Taylor diagram comparing the future common simulated regimes, determined with K-Means clustering, for the CMIP6 ensemble for SSP5-8.5 and the period 2070–2099 to the corresponding reference regimes from ERA5 reanalysis in the historical period (1985–2014), for the Arctic region.

L329: "shifted positive pressure anomaly", comment "Can we learn something from this difference?"

> The discussion will be updated in light of the new results.

Fig. 8: "I would say here there is less of a match."
    Yes, this will be updated.

Fig. 10, "Storylines", comment "This is the average over two model? Can you show them separately?"
    We previously used this figure to back up one claim which is no longer valid in the new and updated results. Therefore we will remove this figure. Since we also include all four storylines we can now refer to the full figures in Levine et al. instead.

L372: "Considered also looking at persistence?"
    We have added analysis of persistence. See our answer to comment 7.

L386: "limited influence of localized drivers", comment "I think it's hard to conclude this based on two models per storyline." and multiple comments from L417 to L420.
    See discussion in answer to comment 9.

L406: "now always...." (probably meant to be "not always...."
    See our answer to comment 5.

L425: "many different considerations", comment "Add references"
    We will add more references to this.

L429: "previously not so well-studied" -> "not often studied"
    We will use this suggestion in the updated manuscript.

Fig. B1: "By eye, this really isn't too clear."
    Correct, but that is the point of the tool, to be able to assist in finding such a point when it is not easy to find by eye.

L460: "smallest area", comment "Why smallest?"
    We want to have an area focused on the Arctic that is not constrained by geometry (due to too small area). If we would extend further south a larger fraction of the domain would be in the mid-latitudes. We will improve this formulation in the revised manuscript.

L460: "The patterns look different in ways, so what is the information that you ate after?"
    At the centre of the domains, the patterns are very similar.

Reviewer 2:

If I were convinced that indeed the projected approach is valid in this case, the article is very well presented and presents a significant contribution to the field. The authors provide an explanaition to why the cluster frequency changes under one storyline and the other, which is a significant contrubution to understanding the impacts of large scale ciruclation changes and the model uncertainty. However, the article would highly benefit from carefully confirming that the regimes for which the frequencies are studied are indeed representative of how the circulation is organized in models and in the future. I recommend major revisions but I highly encourage the authors to address and resubmit.

Major comments:
1. The projected approach to weather regimes is an established methodology. However, the paper would be a much more valuable contribution if the authors were able to provide a more robust validation to justify the approach.
- The pseudo-PCs are already a projection of the data onto the EOFs from ERA5, which is already an great assumption. I would suggest that the projected approach was compared with the complete clustering approach performed with the models. To my understanding, only if the circulation in the model can be considered as representative of the reference, the two can be comparable. Formally, this means that the model sub-space is contained in the reference space. One way of quantifying this is the "quantization error" proposed by Quagraine et al. (2020).
- Currently the evaluation is performed using Taylor diagrams and the statistical significance of the changes in the frequency of occurrence under the influence of GHG is performed with a Welch's test. However, these metrics are valid once the above has been tested (i.e. that the model subspace is contained in the reference subspace).

We follow the suggestion of reviewer 2, and use the "quantization error" proposed by Quagraine et al. (2020) to quantify, if the model sub-space is contained in the reference space. The "quantization error" is defined as the mean error of each day's pattern with respect to the reference cluster pattern to which it belongs (Quagraine et al., 2020). As explained in Quagraine et al. (2020), by comparing the error mapping of the GCMs to that of the ERA5 data, one can assess whether the GCM patterns have a greater error than the reanalysis. In this case, the GCM sub-space is not contained in the reference space spanned by the reanalysis data. We also use this argumentation in our answer to reviewer 1, comment 1.

Figure R21 shows the quantization error for the North-Atlantic-Eurasian regimes for ERA5 (last row) and the historical simulations from 1985-2014. Except for NAO+ respectively NAO- for only 2 respectively 3 models the GCM patterns have smaller errors than the ERA5 daily patterns, which justifies the projection approach. A similar conclusion can be drawn for the Arctic regimes. The majority of the GCMs have smaller errors than the ERA5 daily patterns,

except 3 models for the AO+ regime, 3 models for the AO- regime, and one model for the Cluster 5 regime, refer to Figure R22.

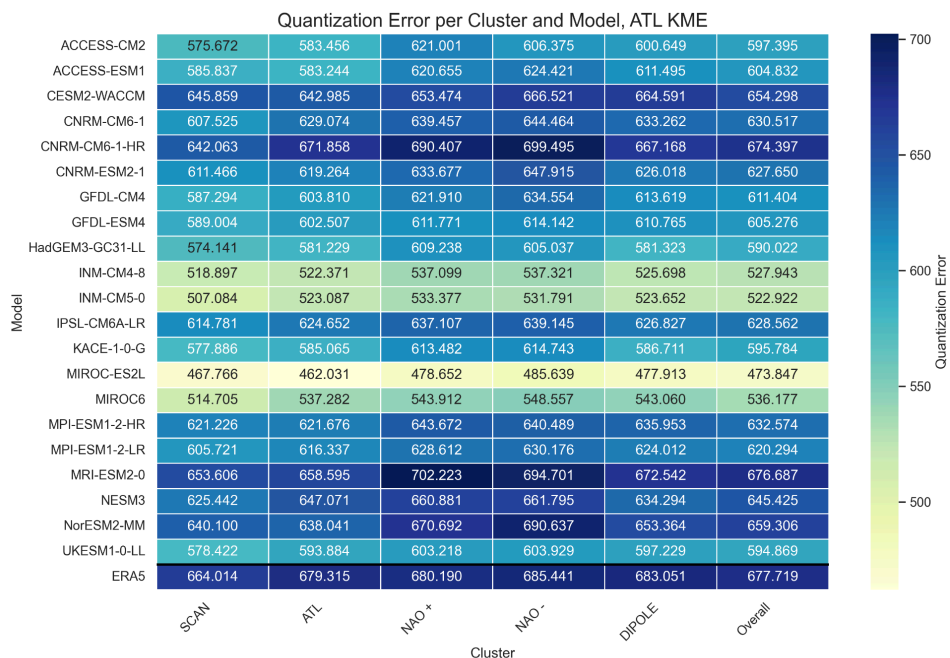



Figure R21 (same as Figure R1): Figure R1: Quantization error per regime and model, North-Atlantic-Eurasian region, KME method, historical period

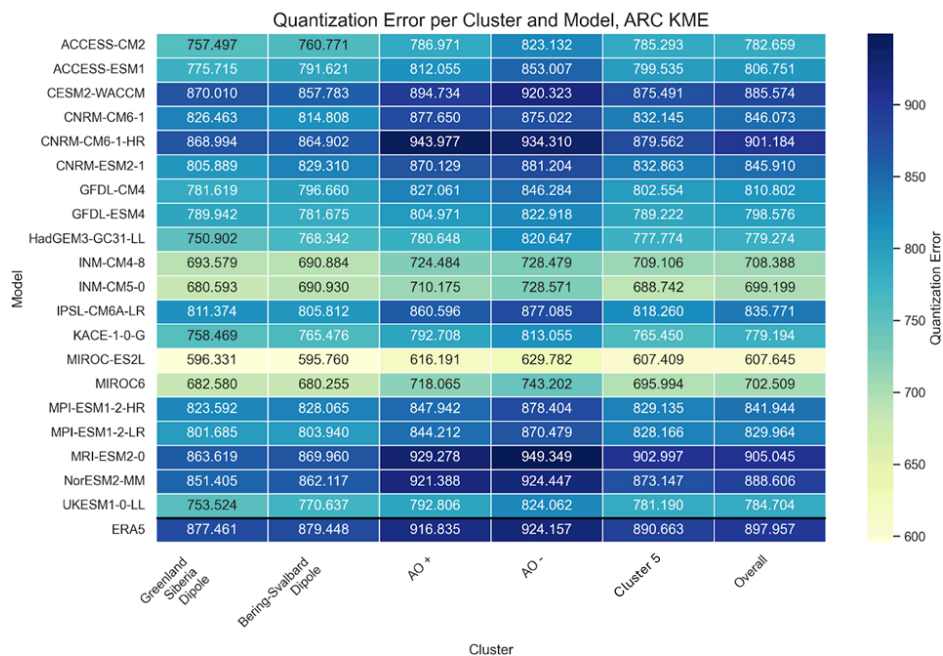



Figure R22 (same as Figure R2): As Figure R20, but for the Arctic region

2. Storylines:
- Why didn't you consider the strong ArcAmp and strong Barents-Kara-Sea warming storyline and the opposite? I can appreciate from Lavine et al. that depending on the target variable, the change under each storyline is different, why use these two only?

As the reviewer mentioned, Levine et al. (2024) produced four Arctic climate change storylines: ArcAmp−/BKWarm+ (A), ArcAmp+/BKWarm+ (B), ArcAmp−/BKWarm− (C), and ArcAmp+/ BKWarm− (D) using the regression approach of Zappa and Shepherd (2017). These storylines show noticeably different paths for Arctic climate change, which deviate substantially from the multi-model ensemble mean (MMM). Analysing these storylines of climate change for different target variables like 2m temperature, u@850hPa, precipitation or sea ice cover in Levine et al (2024) revealed that the storylines A and D are the most contrasting storylines. That's why we considered only these two in the previous version of the manuscript. In the revised version, we consider all 4 storylines.

- Can you show that the circulation in the models undergoes the circulation changes that you consider relevant to your regime changes?

Our circulation regimes are related to specific patterns of jet anomalies (see Figures R15, R16). The CMIP6 multi-model ensemble shows a poleward shifted North Atlantic jet (e.g. Levine et al, 2024,Figure 4, Harvey et al., 2020). Since NAO+ and Dipole are related to northward shifted North Atlantic jet, and NAO- to a southward shifted North Atlantic jet, the robust more frequent occurrence of the NAO+ and Dipole and the less frequent occurrence of NAO- in the future (Fig. R17) are in line with the projected jet changes.

Please refer also to our more in-depth response to reviewer 1, comment 10.

- I think that the assumptions behind using the storylines should be justified. Why can the authors expect to find different regimes of regime frequency under these two storylines?

As shown in Levine et al. (2024, fig. 4), the CMIP6 MMM shows a poleward shifted North Atlantic jet, but the strength of this poleward shift depends on the storyline, with a weaker poleward shift in storyline A and B and slightly stronger poleward shift in storyline C and D (compared to the MMM). Since the circulation regimes are related to specific jet anomaly patterns, one might hypothesize that this is related to future changes in the frequency of occurrence of these regimes, particularly those related to a northward-shifted jet (NAO+, Dipole). We are aware that testing this hypothesis is limited because there are only a few models available which represent each storyline. Please refer also to our response to reviewer 1, comment 9.)

3. Different clustering algorithms. I would move the evaluation of regimes using SAN to the Supporting Information or further exploring the results that are obtained when using SAN for the complete analysis. The authors claim that these is not much different between one method or

the other, however, I find this questionable: First, the correlations in Figure 3 are not impressive and second, the SCAN and DIPOLE regimes are very different from a dynamical perspective. Low slp and high slp over the British Isles respectively can lead to very different weather conditions. What is relevant related to the weather regimes are the regional impacts. At least this is what is claimed in the introduction. Can you show that the precipitaiton and temperature impacts associated with these regimes are comparable?

We argue that it is a strength to have both methods in the paper, as it increases the robustness of the results, so we would prefer to keep it in the main text rather than in an appendix.

As written above, we have updated the results as we discovered an error in our SANDRA setup, improving the agreement between the two methods, justifying that the results are more similar (except Cluster 5 for the Arctic region). This is resembled by the respective composites of other variables like the zonal wind at 250hPa shown in Figures R13-R16 and the composites for 2m air temperature and precipitation shown here in Figures R23 and R24. The zonal wind composites as well as the temperature and precipitation composites are very similar for the two methods, except for Arctic Cluster 5, which has a different circulation, as discussed earlier.

For the composites, we will include the zonal wind composites in the main text and the temperature and precipitation composites in the appendix and add respective discussions.

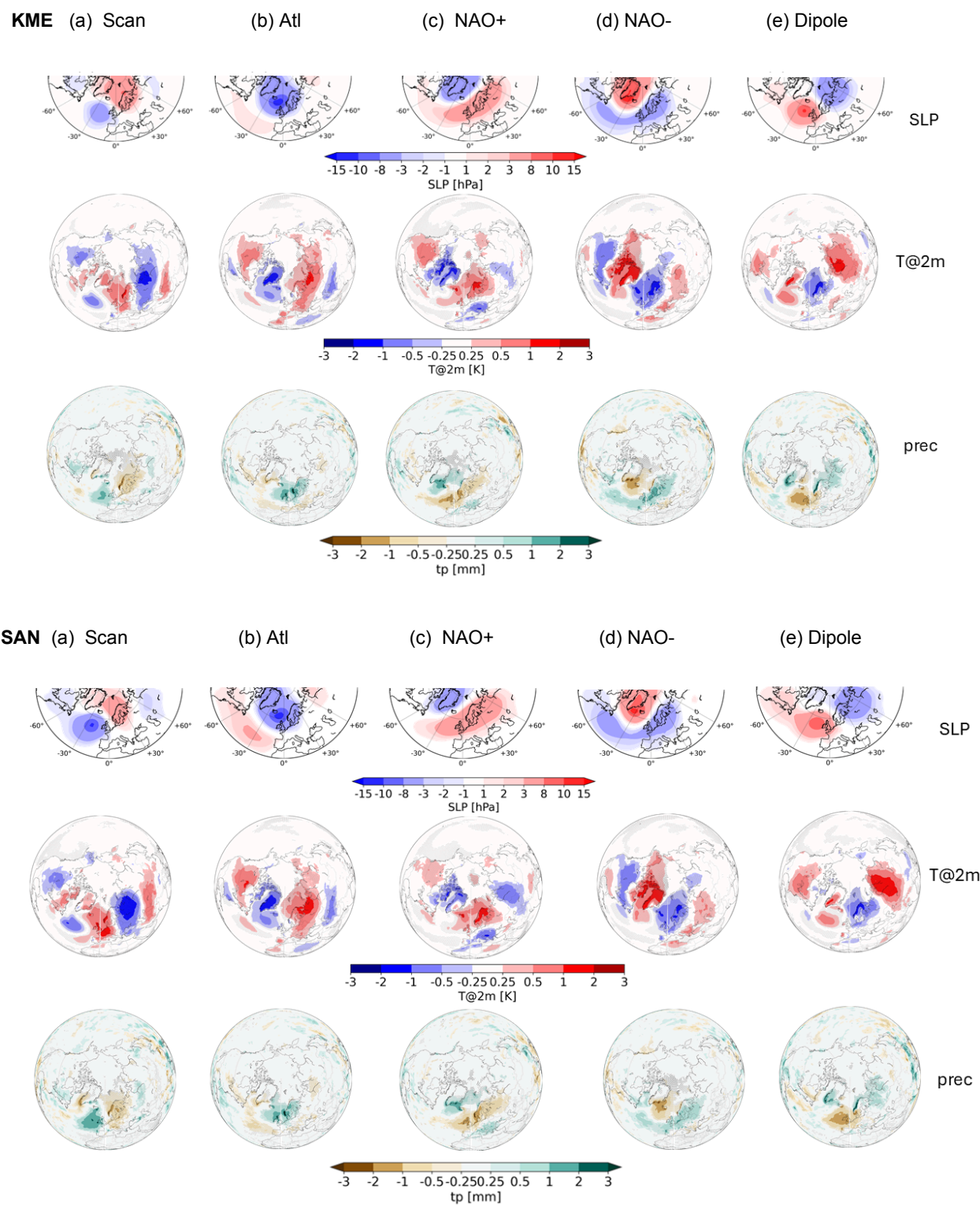



Figure R23: 2m air temperature and total precipitation anomalies associated with the respective regimes for KME and SAN in the North Atlantic-Eurasian region.

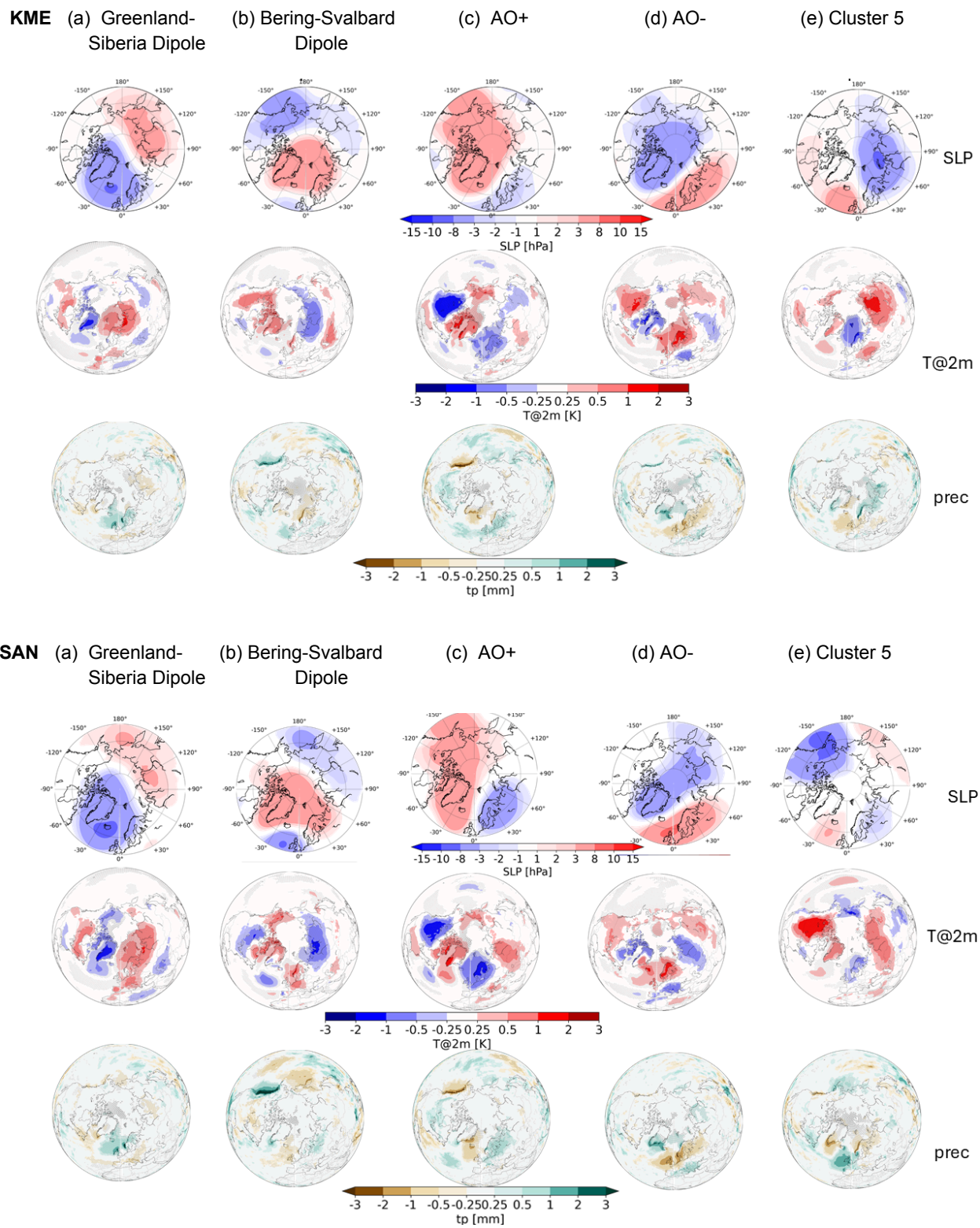



Fig R24: same as R23 but for the Arctic region.

4. I am not convinced by the claim "Thus, the spatial structure of the summer circulation regimes does not change significantly under the influence of rising GHG emission in the future time period compared to the historical time period". Referring to my comment above, I would consider that evaluating the regimes independently for the future and the historical period would be a first step to ensure that projecting the forced simulations is a valid assumption.

> We have argued in our answers to reviewer 1, comments 1 & 2 and reviewer 2 comment 1 that the projected approach is justified. Therefore, we kept the projected approach. To evaluate the future changes in the spatial structure of the circulation regimes we compare the future common simulated regimes for the CMIP6 ensemble with the ERA5 reference regimes for the historical period 1985-2014.

> The respective Taylor plots are shown in Fig. R10 (Fig. 4 in the old manuscript) for the North Atlantic-Eurasian regimes and Fig. R20 (new figure in revised Manuscript) for the Arctic regimes. These Taylor plots underline the similar spatial structure of the common future CMIP6 regimes for SSP5-8.5 and the period 2070–2099 to that of the ERA5 regimes for the historical period 1985-2014, with all pattern correlation coefficients >~0.8. In our view this justifies the application of the "projected approach" (see our terminology described in Fig. 1 of the manuscript) for the calculation of future changes in regime occurrence frequency (and persistence).

> To better describe the calculation of the common simulated regimes we will change the text in the revised manuscript (L122-138) to:

> "The common simulated circulation regime framework enables the possibility to compare the spatial structure between reanalysis and the entire CMIP6 model ensemble. First the preprocessed data of each climate model are merged into a single data file along the temporal axis, either for the historical period 1985-2014 or the future period 2070-2099. Instead of performing a common EOF analysis and obtaining common PCs (as described e.g. in Benestad et al., 2023), here the common climate model data are projected onto the ten-dimensional reference state space, determined from the ERA5 data, resulting in ten common Pseudo-PCs. The common Pseudo-PCs serve as input data for the K-Means clustering algorithm. Five common simulated circulation regimes are obtained for each time period, representing the joint regimes for the entire model ensemble."

If this is the case, then one can claim that the spatial structure of the regimes does not change in the models. I also consider that this is not a "weak external forcing". Given that SSP5-8.5 is used and the authors refer to this as a strong forcing in a previous pragraph.

> We agree that there is a bit of confusion about weak and strong forcing in the manuscript which we clarify here:

> In the conceptual framework of circulation regimes, Palmer (1993, 1999) introduced a dynamical paradigm for climate change. According to this paradigm, a not too strong external forcing does not change the structure and number of atmospheric regimes, but instead changes the frequency of occurrence of the regimes. This determines, at least partly, the time-mean response of the atmospheric flow to the external forcing. On the

other hand, very strong external forcing factors can lead to changes in the number and structure of circulation regimes as proven in several studies, e.g. by Kageyama et al. (1999) and Handorf et al. (2009). Since we have proven that the structure of the circulation regimes is not changing under SSP5-8.5, we call this a weak forcing in the conceptual regime framework and the above mentioned paradigm for climate change.

In the revised manuscript, we will clarify this and avoid the confusion. We will reformulate text beginning L73 (see also L73, reviewer 1 comment) from "In examining the simulation of future time periods, the SSP5-8.5 emission scenario is of particular interest. As the highest emission scenario, it is characterised by a large increase in Greenhouse Gas (GHG) emissions, impinging a strong external forcing upon the atmosphere. Thus, a response of characteristics of circulation regimes is expected." to "For the future period we use the highest emission scenario SSP5-8.5 as it is expected to give the most statistically robust response to climate change."

Moreover, we will reformulate text from L250-254 from "These results support the hypothesis put forth in Section 1, namely that a weak external forcing acting on the dynamical system—in this case, the external forcing on the atmosphere—does not alter the spatial structure of the regime patterns. Thus, the spatial structure of the summer circulation regimes does not change significantly under the influence of rising GHG emission in the future time period compared to the historical time period." to "These results support the hypothesis of Palmer (1993, 1999) introduced in Section 1, namely that a rather weak external forcing acting on the dynamical system atmosphere does not alter the spatial structure of the regime patterns. In our case, even under the high emission scenario SSP5-8.5 the spatial structure of the summer circulation regimes does not change significantly in the future time period."

Line 280: This first sentence is not clear. Do you mean that "In the models representing the BK+ and PA- storylines the NAO+ occurs significantly more?

We are aware of the limitation of the low number of models per storyline, and we cannot prove a consistent response. Nevertheless, it may still be useful to keep the labeling just to indicate that models within the same storyline may actually exhibit quite different frequency responses. In relation to the changes in jet structure (Figures R13-R16), it can for example be that while there is a clear response in the mean jet shift, individual models among storylines project these changes on different clusters.

Line 295: The immediate above paragraphs were describing the results in Figure 6, while the paragraph starting in l288 refers to the KAN and SAN comparison. I don't follow the logic.

L295 is also part of the summary (starting with L288) and refers to results from both KME and SAN.

Caption Figure 6: The markers associated with the storylines are not escribed.

Label descriptions will be added to the figure caption. (See updated Figures R11 & R12 for persistence or R17 & R18 for frequency of occurrence.)

Overall the silhouette and elbow results are not impressive, 6 or 7 could have been a perfectly reasonable number of clusters. Line 440 says "The KElbow Visualizer suggests that five clusters represent an optimal number of clusters for both regions." No description of what is seen in the figure is provided. Why is this suggested by the figure?

We agree, that "optimal" number of clusters is an overstatement. We will reformulate from optimal to reasonable/suitable number in the revised manuscript. In addition, we will improve the description of the results from the elbow plots shown in the appendix. The elbow plot displays the distortion score, which describes the within-cluster sum of squared distances, in dependence of the number of clusters. The point of greatest curvature of that curve, the so-called "elbow" point suggests a reasonable number of clusters. As we agree that the change in the curvature is not very strong from k=4 to k=6 clusters, we justify our choice of 5 clusters with two other metrics (in the appendix). This is the silhouette score (already included in the old manuscript) and an additional metric, following Grams et al. (2017), a reference reviewer 1 suggested in comment 8. As suggested by Grams et al (2017) their metric to evaluate a suitable number of clusters is the Anomaly Correlation Coefficient ("*The optimal number of clusters is seven (Supplementary Fig. 1) based on the criterion that the anomaly correlation coefficient between the clusters is below 0.4*", Grams et al., 2017, Method section). In the answer to reviewer 1, comment 3 we show the results of the anomaly correlation coefficient calculated for the North Atlantic-Eurasian region in Figure R6 and for the Arctic region in Figure R7 for both methods, KME and SAN, based on ERA5 data 1985-2014. For both methods and both regions, the suggested criteria of Grams et al. (2017) is satisfied for 5 regimes each, which justifies our choices for the number of regimes.

Comments from the PDF:

L40: This paragraph, and in particular the second sentence are not. The logic is not clear. First sentence says that detection of circulation *changes* is challenge due to signal-to-noise problem. How is this related to evaluating state-of-the-art simulation of present day regimes?

We will reformulate L39-43 into:

"A reliable detection of circulation changes in climate model simulations is often limited by a large uncertainty and low signal-to-noise ratios (Scaife and Smith, 2018; Smith et al., 2022). Nonetheless, the concept of atmospheric circulation regimes have been successfully applied to characterise future circulation changes (e.g. Fabiano et al. 2021). A first step is to evaluate the ability of the models to reproduce observed circulation Regimes. To get robust results for future changes in circulation regimes, one needs to consider multi-model ensembles of climate projections such as the Coupled Model Intercomparison Project (CMIP)."

L69: "physically based" -> "dynamical"?

The derivation in Levine et al. 2024 is based not on dynamical changes but on physical criteria (target variables relevant for climate change risks), so we would prefer to keep this formulation.

L77: "strong external forcing" "below you say it is a weak external forcing... are you referring to a different forcing?"

> We will remove "strong" from this sentence. It was originally meaning "strong" as it is a high emission scenario, but we see that it can be confusing. Also refer to our answer to reviewer 2, comment 4, second part.

L80: Regarding "since the circulation regimes in this area are well-known." comment:
"Can you provide a reference for this? I think that I understand what you mean with "well known", but if the reference to previous research is going to be the only justification for the choice of region you should at least state it formally" "Given that the circulation regimes in this area have been extensively studied (provide corresponding references), and [the a reason to why this is enough to select a region...]."

> We take the reviewer's suggestion and use:
> "Given that the circulation regimes in this area have been studied before (Crasemann et al. 2017) during winter, we now want to study this area in the summer season."

L82: "this" -> "region selection"

> We will use this suggestion in the updated manuscript.

L92: Comment on "Averaging the data day by day at each grid point yield the mean seasonal cycle.": "Isn't this the definition of a daily climatologu? Why not just say that you substract the daily climatology".

> Yes, it is, but we wanted to be explicit.

L92-93: Comment on "Additionally, a 21-day running mean is applied to remove higher-frequency fluctuations": "If daily weather regimes are the focus of the study, wouldn't this filter remote part of the signal under study?"

> We are sorry for the misunderstanding. We only perform the running mean on the climatology that we subtract. This keeps the high frequency changes in the resulting data. We will rewrite this to make this clear in the updated manuscript. We are replacing "Averaging the data day by day at each grid point yield the mean seasonal cycle. Additionally, a 21-day running mean is applied to remove higher-frequency fluctuations" by "The mean seasonal cycle is obtained by computing the daily climatology and then applying a 21-day moving average to this.".

L122 and L148 on the methodology:

> These comments are covered by previous answers. Please refer to Reviewer 2, comment 1, and Reviewer 1, comment 3.

Table 1: "I found it confusing that you call it Polar Amplification and they call it Arctic Amplification. It is a detail but it is still confusing to me. "

> We will call it Arctic Amplification (AA for short) in the revised manuscript.

Multiple comments on L230

    These comments are answered in our replies to Reviewer 1, comment 5 & 10, and Reviewer 2, comment 3.

Fig. 4 caption: "Clarify period".

    We will clarify the time period.

L250: "Are you referring to the anthropogenic CO2 forcing? I don't know if I would call this a weak forcing, given the significant circulation changes associated with this forcing."

    Please, refer to Reviewer 2 comment 4.

Comments on L251 and L254:

    Please, refer to Reviewer 1, comment 1 and reviewer 2, comment 1.

L258: "I don't understand. The regimes could be different and therefore the frequency could be different. If by construction you force the regime centroids to be different to what the true centroids are in the models, how can you be sure that these frequencies are representative of how the actual circulation states?."

    Please, refer to Reviewer 1, comment 4.

New references:

Barnes, E. A., and L. Polvani, 2013: Response of the Midlatitude Jets, and of Their Variability, to Increased Greenhouse Gases in the CMIP5 Models. J. Climate, 26, 7117–7135, https://doi.org/10.1175/JCLI-D-12-00536.1.

Dawson A, Palmer T (2015) Simulating weather regimes: Impact of model resolution and stochastic parameterization. Clim Dyn 44(7–8):2177–2193. https://doi.org/10.1007/s00382-014-2238-x

Fabiano, F., Christensen, H. M., Strommen, K., Athanasiadis, P., Baker, A., Schiemann, R., & Corti, S. (2020). Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance. Climate Dynamics, 54(11), 5031-5048. https://doi.org/10.1007/s00382-020-05271-w

Falkena SK, de Wiljes J, Weisheimer A, Shepherd TG. Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector. Q J R Meteorol Soc. 2020; 146: 2801–2814. https://doi.org/10.1002/qj.3818

Grams, C., Beerli, R., Pfenninger, S. et al. Balancing Europe's wind-power output through spatial deployment informed by weather regimes. Nature Clim Change 7, 557–562 (2017). https://doi.org/10.1038/nclimate3338

Harvey, B. J., Cook, P., Shaffrey, L. C., & Schiemann, R. (2020). 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(23), e2020JD032701. https://doi.org/10.1029/2020JD032701

Michelangeli, P. A., Vautard, R., & Legras, B. (1995). Weather regimes: Recurrence and quasi stationarity. Journal of Atmospheric Sciences, 52(8), 1237-1256. https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2

Philipp, A., Della-Marta, P. M., Jacobeit, J., Fereday, D. R., Jones, P. D., Moberg, A., & Wanner, H. (2007). Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. Journal of Climate, 20(16), 4065-4095. https://doi.org/10.1175/JCLI4175.1

Quagraine, K. A., Nkrumah, F., Klein, C., Klutse, N. A. B., & Quagraine, K. T. (2020). West African Summer Monsoon Precipitation Variability as Represented by Reanalysis Datasets. Climate, 8, Article 111. https://doi.org/10.3390/cli8100111

Skific, N., Francis, J. A., & Cassano, J. J. (2009). Attribution of projected changes in atmospheric moisture transport in the Arctic: A self-organizing map perspective. Journal of Climate, 22(15), 4135-4153. https://doi.org/10.1175/2009JCLI2645.1

Vautard R. 1990. Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. Mon. Weather Rev. 118: 2056–2080. https://doi.org/10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;2

*Woollings, T., and M. Blackburn, 2012: The North Atlantic Jet Stream under Climate Change and Its Relation to the NAO and EA Patterns. J. Climate, **25**, 886–902, https://doi.org/10.1175/JCLI-D-11-00087.1.*