Response to Reviewer Comments

A low-dimensional framework for interpreting Northern Hemisphere winter extratropical precipitation trends

egusphere-2026-3431 (WCD)

We thank the reviewer for a careful, constructive reading of our manuscript. Below, we respond to each comment in turn (reviewer text in *italics*) and show how we have revised the manuscript.

[Reviewer 2]

Major Comments

Major Comment 1

> *I have concerns about the interpretations regarding the use of zonal mean geopotential height to separate circulation or thermodynamic-driven changes. An example of this is in the abstract, where the authors say circulation-induced precipitation trends become weaker relative to total precipitation trends. However, in reanalysis trends, the zonal mean component typically has a small contribution to the overall pattern of Z500. Whereas in the models, Z500 changes are typically more uniform and therefore the zonal mean Z500 has a large contribution to the overall pattern. This complicates the interpretations regarding the decomposition that are being applied uniformly across products. We will need to see spatial maps of Z500 to gauge if this interpretation is appropriate. Some references that may address this are Wettstein & Wallace (2010), Wills & Schneider (2016), or Wills et al (2016), which discuss the non-zonal component that is lost in this analysis. This is particularly confounding over the North Atlantic, where the jet and storm track over the period analyzed here shifted equatorward, which is opposite of the Pacific. These issues of separating dynamics and thermodynamics as well as the loss of zonally asymmetry through the approach are discussed briefly in the discussion but could be addressed more thoroughly throughout the rest of the manuscript.*

Response:

We thank the reviewer for raising these concerns about the spatial information retained by the framework and the interpretation of the reconstructed precipitation trends. We first clarify that our EOF analysis is applied to zonal-mean geopotential-height activity, rather than to zonal-mean geopotential height itself. The activity is calculated from squared height fluctuations at each latitude–longitude grid point before zonal averaging. Consequently, zonally asymmetric disturbances contribute to the activity measure even when their height anomalies cancel in the zonal mean. Therefore, the relative size of the zonal-mean component of the Z500 trend, which differs between reanalysis and models, does not by itself indicate how much variability our activity-based reconstruction represents. Nevertheless, this distinction does not imply that all regional circulation variability is captured.

Our objective is to diagnose Northern Hemisphere meridional precipitation redistribution using a compact set of circulation-activity modes. For this objective, cancellation between opposing regional precipitation trends is part of the zonal-mean quantity being examined. Regarding the opposite jet and storm-track shifts in the two basins noted by the reviewer, the framework does not require the North Pacific and North Atlantic to exhibit identical circulation changes. Moreover, precipitation is regressed onto the PCs at each latitude–longitude grid point, allowing the associated precipitation patterns to differ between basins before their zonal means are calculated.

We agree, however, that regional circulation variability not represented by the retained PCs can be omitted, and that the importance of this omission may differ between reanalysis and models. In particular, opposing activity changes between basins can have a weak zonal-mean signature while still contributing to zonal-mean precipitation if their regional precipitation sensitivities differ. We have clarified this limitation and the distinction between hemispheric and regional interpretations. The relevance of sectoral storm-track variability and zonally asymmetric hydrological responses is now discussed with reference to Wettstein and Wallace (2010), Wills and Schneider (2016), and Wills et al. (2016).

To make the spatial structures of circulation variability explicit, we have added Supplementary Fig. 5, which shows the spatial patterns of precipitation regressed onto the leading three PCs derived from the EOF analysis of zonal-mean transient Z500 activity for ERA5, CMIP6 historical, and SSP5-8.5 simulations. Because the regression is performed at each latitude–longitude grid point, these maps show how each zonal-mean mode is expressed regionally, including over the North Pacific and North Atlantic. As a complementary analysis, we also performed EOF analysis directly on two-dimensional latitude–longitude fields of transient Z500 variance. The corresponding circulation activity and precipitation regression patterns for ERA5 are presented in Fig. R2 (also shown in our response to Reviewer 1). Unlike the zonal-mean modes, these two-dimensional EOFs partition variability largely by basin (e.g., EOF1 over the North Atlantic and EOF3 over the North Pacific in ERA5), so the hemispheric meridional signal is distributed among regional modes whose ordering and interpretation may not be consistent across datasets. These diagnostics complement the zonal-mean analysis but are not intended to represent the full regional circulation variability or the complete Z500 trends.

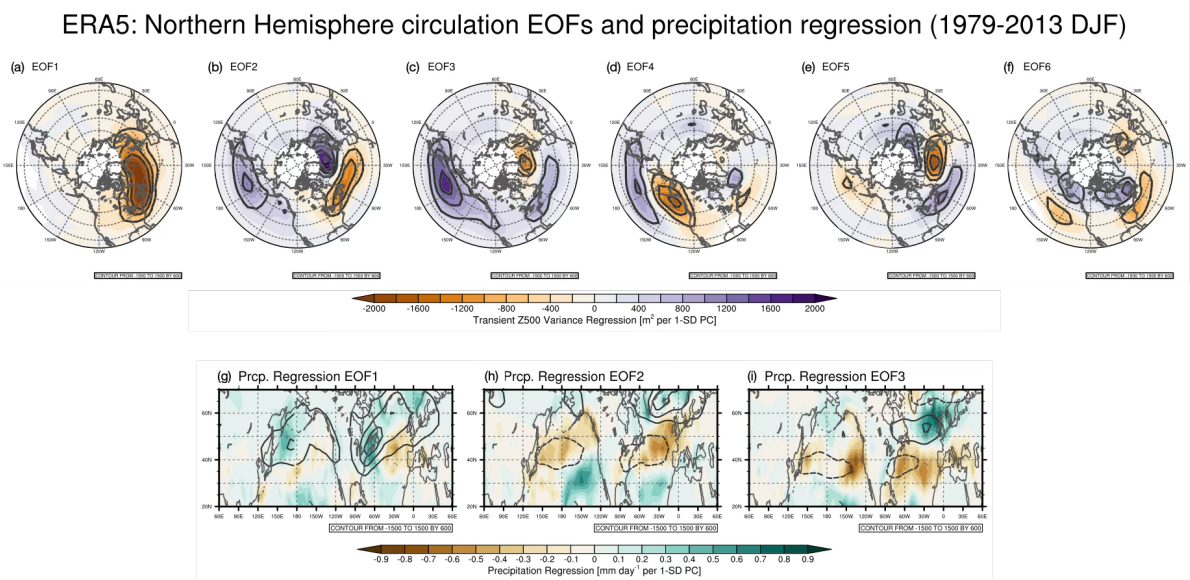


Figure R2. Spatial regression pattern of Northern Hemisphere circulation activity and precipitation associated with the leading two-dimensional EOF modes in ERA5. (a–f) Latitude–longitude distributions of transient Z500 variance regressed onto the standard principal components (PCs) of EOF1–EOF6, respectively. (g–i) Corresponding precipitation regression patterns for EOF1–EOF3. The EOF modes were derived directly from two-dimensional latitude–longitude fields of transient Z500 variance during boreal winter (DJF, 1979–2013), retaining both latitudinal and longitudinal variability. Shading indicates regression coefficients for transient Z500 variance (m^2 per 1-SD PC) and precipitation (mm day^{-1} per 1-SD PC), respectively.

We also agree that the reconstruction should not be interpreted as a complete separation of dynamical and thermodynamic precipitation changes. It diagnoses precipitation changes linearly associated with the retained circulation-activity modes. These statistical relationships can depend on the moisture and thermal background, and the unreconstructed trend can include omitted circulation variability as well as thermodynamic effects and other processes. Although this limitation was acknowledged in the original discussion, some statements in the abstract and results implied a stronger attribution than the framework supports.

We have therefore (1) removed the inference that a smaller relative reconstruction demonstrates an increasing thermodynamic contribution, (2) revised the terminology to emphasize circulation-associated precipitation changes, and (3) clarified that reconstruction performance evaluates the retained statistical representation rather than the full dynamical contribution. In particular, the revised SSP5-8.5 results are described as capturing the broad meridional precipitation-trend pattern while underestimating its amplitude. These changes preserve the intended hemispheric application of the framework while making its regional and attribution limits explicit.

Major Comment 2

> *I have concerns about the sampling of models between the historical and future scenarios. I know it is difficult to get the same model outputs based on what has been made available. However, the interpretations and their utility going forward for model development and evaluation are complicated by this analysis. It would be helpful in the scatter plots to see which models lie in which High/Low groups by identifying them individually.*

Response:

We thank the reviewer for raising this concern. We agree that differences in model availability between the historical and SSP5-8.5 experiments may introduce sampling uncertainty into their comparison. We also agree that the identities of individual models and the members of the selected model groups should be clearly reported.

To assess the potential influence of differences in model sampling, we repeated the analysis (Fig. R1) using the same 12 models available for both the historical and SSP5-8.5 experiments: CanESM5, INM-CM4-8, ACCESS-CM2, CESM2-WACCM, EC-Earth3, IPSL-CM6A-LR, MIROC6, MRI-ESM2-0, NorESM2-LM, NorESM2-MM, TaiESM1, and KACE-1-0-G.

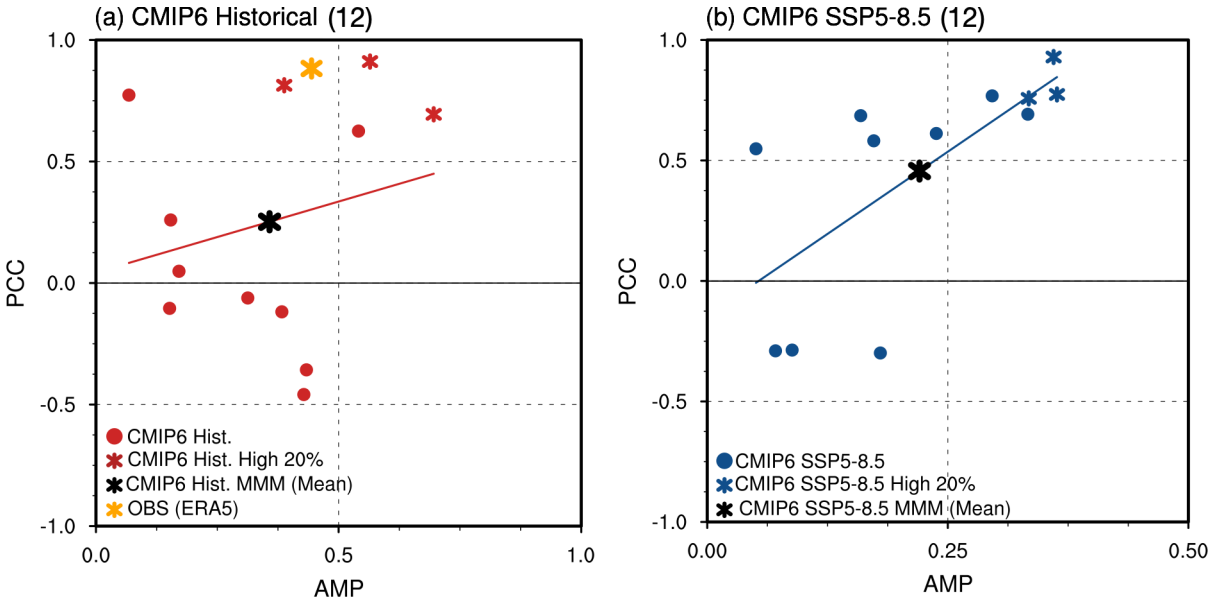


Figure R1. PCC–AMP relationship for circulation-based precipitation reconstructions using the 12 CMIP6 models common to the historical and SSP5-8.5 simulations. Results are shown for (a) historical and (b) SSP5-8.5 simulations. Colored circles indicate individual CMIP6 models, colored asterisks indicate the three models with the lowest NRMSE values (25% of the common ensemble), and black asterisks denote the arithmetic means of the individual-model PCC and AMP values. The orange asterisk in (a) indicates ERA5. Solid lines show the fitted linear relationships between PCC and AMP across models. The three selected models are MRI-ESM2-0, ACCESS-CM2, and KACE-1-0-G for (a), and NorESM2-LM, MRI-

ESM2-0, and CESM2-WACCM for (b).

The results obtained from the common 12-model ensemble were broadly consistent with those based on the full ensembles (21 historical and 19 SSP5-8.5 models). Specifically, the regression slopes between PCC and AMP changed from 0.61 to 0.58 (a 3.8% decrease) in the historical simulations and from 2.95 to 2.73 (a 7.5% decrease) in the SSP5-8.5 simulations. The corresponding means of the individual-model PCC values changed from 0.32 to 0.25 in the historical simulations and from 0.45 to 0.46 in the SSP5-8.5 simulations. These results indicate that the overall PCC–AMP relationships remain similar when the same 12 models are used for both experiments, suggesting that the main model-evaluation findings are not primarily driven by differences in model sampling. The comparison is presented in Fig. R1.

Supplementary Table 2. CMIP6 models selected for the high- and low-skill subsets of each EOF mode in the historical and SSP5-8.5 simulations. For each EOF mode, the five models with the lowest and highest NRMSE values are classified as high- and low-skill models, respectively. The selected subsets correspond to those shown in Fig. 5(a)–(f).

Simulation	Historical Simulation		SSP5-8.5 Simulation	
EOF/Skill	High-skill group	Low-skill group	High-skill group	Low-skill group
EOF1	GFDL-CM4	CESM2-WACCM	NorESM2-LM	MIROC6
	NorESM2-LM	CESM2	MRI-ESM2-0	ACCESS-CM2
	ACCESS-CM2	EC-Earth3	NorESM2-MM	AWI-CM-1-1-MR
	MRI-ESM2-0	FGOALS-g3	CESM2-WACCM	INM-CM4-8
	MPI-ESM-1-2-HAM	CESM2-FV2	TaiESM1	EC-Earth3-CC
EOF2	MPI-ESM1-2-LR	INM-CM4-8	EC-Earth3-Veg-LR	NorESM2-LM
	KACE-1-0-G	IPSL-CM6A-LR	EC-Earth3	CanESM5
	CESM2-FV2	MIROC6	EC-Earth3-Veg	MRI-ESM2-0
	ACCESS-CM2	NorESM2-MM	CESM2-WACCM	NorESM2-MM
	CESM2	NorESM2-LM	MIROC6	KACE-1-0-G
EOF3	MPI-ESM-1-2-HAM	CanESM5	CanESM5	MIROC6
	NorESM2-MM	INM-CM4-8	NorESM2-LM	BCC-CSM2-MR
	MRI-ESM2-0	GFDL-CM4	EC-Earth3-CC	ACCESS-CM2
	MIROC6	EC-Earth3	INM-CM5-0	TaiESM1
	IITM-ESM	MPI-ESM1-2-HR	EC-Earth3	INM-CM4-8

To improve the transparency of model selection, we have added information on the selected high-ranking models to the captions of Figs. 1 and 2 (MRI-ESM2-0, MPI-ESM-1-2-HAM, ACCESS-CM2, and KACE-1-0-G for CMIP6 historical simulations, and NorESM2-LM, MRI-ESM2-0, CESM2-WACCM, and EC-Earth3 for CMIP6 SSP5-8.5 simulations). In addition, we have added Supplementary Table 2, which explicitly lists the models included in the High and Low groups for each EOF mode shown in Fig. 5. Following the revision of the model-evaluation metric, these groups are now defined using NRMSE rather than PROJ.

Major Comment 3

The conclusions stated for the zonal mean temperature differences between groups do not seem well supported by Figure 8. In particular similarities between the historical and SSP585 scenario are difficult to see. There seems to be a consistent signal at low-levels over the Arctic, but the other aspects mentioned do not clearly convey from the figure. Perhaps it is more in their differences that the authors would like to explore more. A re-evaluation of this section is needed.

Response:

We thank the reviewer for this helpful comment. After re-examining Fig. 8, we agree that the original text overstated the physical interpretation of the high-minus-low-skill temperature-trend differences and gave too little attention to how the two periods differ. We have revised Section 3.3 accordingly.

First, we now state explicitly how the historical and SSP5-8.5 simulations differ. The difference in the subtropical upper troposphere changes sign between the two periods. In the historical simulations, high-skill models warm less than low-skill models for EOF3 (Fig. 8c). Under SSP5-8.5, they warm more for both EOF1 and EOF3 (Figs. 8d, f). This contrast is consistent with Chemke and Yuval (2026), cited in the same paragraph of the revised manuscript, who linked enhanced North Pacific winter storm-track activity to subtropical tropospheric warming and a stronger midlatitude meridional temperature gradient.

Second, we have made the physical interpretation of the historical differences more cautious. For EOF1, the enhanced lower-tropospheric meridional temperature-trend gradient is now placed on the poleward flank of the storm track (~50–65°N), where the signal appears in Fig. 8a, rather than near 45°N. Its link to baroclinic growth is now described as possible rather than established. For EOF2, we removed the inference that the temperature-trend differences lead to an equatorward storm-track displacement, because Fig. 8b does not support it. We now state only that the thermal differences may influence the distribution of baroclinity and the associated storm-track response.

As a result, we have made the following changes to the revised manuscript:

L391–393: "This structure implies a stronger lower-tropospheric meridional temperature-trend gradient on the poleward flank of the storm track (~50–65°N), which may favor stronger baroclinic growth and thus enhanced transient-eddy activity and the associated precipitation response."

L395–397: "…together with a weak negative temperature-trend difference near 45°N in the mid-to-lower troposphere, suggesting differences in the meridional thermal structure between the model subsets. These thermal differences may influence the distribution of baroclinity and the associated storm-track response."

L418–420: "In addition, the subtropical upper-tropospheric difference changes sign between the two periods, from weaker warming in high-skill models for EOF3 in the historical simulations (Fig. 8c) to stronger warming for EOF1 and EOF3 under SSP5-8.5 (Figs. 8d, f)."

L439–443: “In particular, the pulsing- and shifting-related modes (EOF1–EOF2) exhibit large dispersion across models, with frequent sign disagreement for EOF1, whereas the sharpening-related mode (EOF3), despite mixed-sign projections in the historical simulations, is more consistently represented under SSP5-8.5. The EOF1 sign disagreement persists under SSP5-8.5, indicating that uncertainty in future precipitation redistribution is closely tied to how models simulate changes in storm-track variability.”

Minor Comments

Minor Comment 1

Line 20-22: This sentence is somewhat vague on what SSP5.85 refers to. Is it part of the historical record? Is it end of century? Trends?

Response:

We thank the reviewer for pointing out this ambiguity. We clarified that SSP5-8.5 refers to future climate projections rather than part of the historical record or an end-of-century climatological difference. The SSP5-8.5 analysis covers the period 2015–2100 and examines linear DJF trends over this period. For consistency with the 35-year historical analysis, the SSP5-8.5 trend slopes are expressed as changes per 35 years.

Minor Comment 2

Line 28-29: But these perspectives do not generally assume fixed relative humidity regionally in the extratropics.

Response:

We thank the reviewer for this clarification. We have revised the sentence as follows:

L31–33: "From a moisture budget perspective, thermodynamic change reflects changes in atmospheric water vapor at fixed circulation (Trenberth et al., 2003; Held and Soden, 2006), whereas dynamic change arises from changes in circulation at fixed atmospheric moisture (Seager et al., 2010)."

Minor Comment 3

Line 58: Blackport & Fyfe (2022) also show this for the wintertime North Atlantic jet and relate these model-observational discrepancies to precipitation.

Response:

We thank the reviewer for suggesting this relevant reference. We added Blackport and Fyfe (2022) to the Introduction and expanded the discussion of observed–model discrepancies in extratropical circulation trends. In particular, we now note that climate models fail to reproduce the observed strengthening of the wintertime North Atlantic jet and that this discrepancy has important implications for precipitation trends over Europe.

The following sentence has been added:

L61–63: "Consistent with these broader circulation biases, Blackport and Fyfe (2022) showed that climate models fail to reproduce the observed strengthening of the wintertime North Atlantic jet, leading to corresponding discrepancies in simulated precipitation trends over Europe."

Minor Comment 4

Line 91-93: The description here is somewhat misleading. Since data availability is limited, the description should be expanded to include that only around 12 models have availability for the historical and future scenarios. A statement on sensitivity to only using the overlapping model outputs would be beneficial as well. Looking at the table of models used, there is much less overlap in the model outputs available in both the historical and SSP585 ensembles. This means the comparisons between the two ensembles has a lot of unaccounted for model uncertainty.

Response:

Please see our response to Major Comment 2 for a detailed description of the common-model sensitivity analysis.

We revised the Data section to explicitly state that the primary analysis uses 21 historical models and 19 SSP5-8.5 models, of which 12 models are common to both experiments. We also clarified that these 12 common models are highlighted by shading in Supplementary Table 1 to facilitate their identification.

As described in our response to Major Comment 2, we repeated the analysis using the common 12-model cohort. The overall PCC−AMP relationships remained broadly consistent with those obtained from the full ensembles, supporting the robustness of the main model-evaluation findings. We therefore retained the full 21- and 19-model ensembles for the primary analysis and presented the common-model sensitivity analysis in Fig. R1 of our response.

Minor Comment 5

Line 126-129: How separated are the three modes using a North (1982) test?

Response:

We thank the reviewer for this suggestion. We applied the North et al. (1982) sampling-error test to determine whether the leading transient EOF modes are distinguishable within their sampling uncertainties. The test was performed for ERA5 and separately for each CMIP6 historical and SSP5-8.5 model. The eigenvalue sampling errors were estimated using effective sample sizes accounting for temporal autocorrelation in the corresponding PC time series.

For ERA5, the sampling-error ranges of EOF1–EOF4 did not overlap between adjacent modes, indicating that EOF1, EOF2, and EOF3 are distinguishable from their respective following modes. The same result was obtained for all 21 historical models and all 19 SSP5-8.5 models based on model-specific North tests. In addition, EOF1–EOF3 collectively explained 88.3% of the transient variability in ERA5, compared with 90.4% and 90.9% on average across the historical and SSP5-8.5 models, respectively.

These results support treating EOF1–EOF3 as separately distinguishable modes rather than as a degenerate EOF subspace. The detailed North-test results, including eigenvalues, sampling errors, and mode-separation assessments, are presented in Table R1 of this response. Although EOF4 is not shown in the table, it was included in the North test to assess the separation between EOF3 and EOF4.

Table R1. North-test results for the leading transient EOF modes. Explained variance and eigenvalue sampling errors are shown for ERA5 and the CMIP6 historical and SSP5-8.5 simulations. CMIP6 values represent the arithmetic means of model-specific results across 21 historical and 19 SSP5-8.5 models, respectively. Sampling errors were estimated following North et al. (1982). Separation was assessed based on whether the sampling-error ranges of adjacent EOF modes overlap.

Values in parentheses indicate the number of models satisfying the separation criterion out of the total number of models. Separation of EOF3 was evaluated against EOF4, which is omitted from the table for brevity.

Dataset	EOF	Explained variance (%)	Eigenvalue ± North error	Separation from next EOF
ERA5	**EOF1**	48.52	1,261,165 ± 178,653	**Yes**
	EOF2	25.67	667,290 ± 91,534	**Yes**
	EOF3	14.09	366,173 ± 47,082	**Yes**
CMIP6 Historical	**EOF1**	54.12	1,361,476 ± 197,543	**Yes (21/21)**
	EOF2	24.56	629,705 ± 87,125	**Yes (21/21)**
	EOF3	11.76	305,156 ± 38,935	**Yes (21/21)**
CMIP6 SSP5-8.5	**EOF1**	56.32	1,761,662 ± 160,409	**Yes (19/19)**
	EOF2	23.25	709,291 ± 62,948	**Yes (19/19)**
	EOF3	11.37	348,016 ± 28,589	**Yes (19/19)**

Minor Comment 6

Line 133: You show 35-yr trends for SSP585, but can you specify what time period you are using or if you are using a pseudo-ensemble type approach? I only see 2015-2100. This should also be clarified in the caption for Figure 1.

Response:

We thank the reviewer for pointing out the ambiguity in the definition of the SSP5-8.5 trends. We have clarified the analysis periods and trend calculations for all datasets. Specifically, ERA5 and CMIP6 historical simulations are analyzed over the 1979–2014 period, whereas SSP5-8.5 simulations are analyzed over the full 2015–2100 period. For each dataset, linear trends are calculated over the corresponding analysis period and expressed as changes per 35 years to ensure consistency in the presentation and comparison of the results. Thus, the SSP5-8.5 trends represent linear trends estimated over the entire 2015–2100 period and subsequently scaled to a 35-year equivalent, rather than trends calculated using moving 35-year windows or a pseudo-ensemble approach.

We have reviewed all figures and their captions in both the main manuscript and Supplementary Material and revised those presenting trends to clearly specify the corresponding analysis periods and 35-year trend units.

Minor Comment 7

Figure 1 caption: Add the precipitation units.

Response:

We thank the reviewer for this suggestion. We have added the precipitation-trend units to the caption of Figure 1.

Minor Comment 8

Line 175-177, 191-192: Looking at Fig. 1 and Supplementary Fig. 2, the differences between the models and reanalysis look more like amplitude differences rather than pattern differences. Depending on how you define your metrics, it seems like the models would still capture a "shift".

Response:

We thank the reviewer for this important observation. We agree that the differences between ERA5 and CMIP6 should be interpreted by distinguishing spatial-pattern agreement from amplitude differences. In particular, a weak reconstructed precipitation trend does not necessarily indicate a failure to reproduce the shift-like spatial pattern, as it may also reflect an underestimated amplitude. We acknowledge that our original interpretation did not sufficiently distinguish these two aspects.

To address this issue, we revised the model-evaluation framework by adopting NRMSE as the primary metric, while retaining PCC and AMP as complementary diagnostics of spatial-pattern agreement and relative amplitude, respectively. PROJ (PCC × AMP) quantifies the reconstructed trend's projection onto the target trend pattern, whereas NRMSE measures reconstruction error arising from pattern mismatch and amplitude underestimation or overestimation. PROJ is retained for comparison with the original analysis and for quantifying the contributions of individual EOF modes. Importantly, although adopting NRMSE changes the rankings of some individual models, the main findings regarding the dominant contribution of transient variability and the substantial intermodel differences in EOF-mode contributions remain unchanged.

L200–201: "Applying the EOF–regression framework to ERA5 reproduces the broad DJF precipitation-trend pattern but underestimates its amplitude (Fig. 1a)."

L214–218: "This weaker reconstructed amplitude in the CMIP6 MMM than in ERA5 indicates that the retained circulation-activity modes account for a smaller fraction of the simulated precipitation-trend magnitude. Although previous studies have reported that climate models tend to underestimate observed poleward storm-track migration (Chemke and Yuval 2026), the amplitude discrepancy identified here does not, by itself, establish a failure to reproduce the shift-like spatial pattern."

L435–439: "Although adopting NRMSE as the primary evaluation metric changes the rankings of some individual models, the main conclusions regarding the dominant contribution of transient variability and the substantial intermodel differences in EOF-mode contributions remain unchanged. The complementary use of PCC and AMP further clarifies that weak reconstructed precipitation trends can reflect amplitude underestimation without necessarily indicating the absence of the corresponding spatial patterns."

We revised the Results section to clarify that the CMIP6 simulations reproduce aspects of the broad shift-like precipitation pattern, although the reconstructed amplitude is generally underestimated. We no longer interpret a weak reconstructed trend as evidence that the shift-like pattern is entirely absent. We also revised the Discussion and Conclusions section to emphasize that the main conclusions remain robust under the NRMSE-based evaluation, while the complementary use of PCC and AMP allows a clearer distinction between spatial-pattern agreement and amplitude bias.

Minor Comment 9

Line 206: The approach uses Z500, which includes a temperature component. When averaged zonally, this impact of temperature should be even stronger.

Response:

We thank the reviewer for this important clarification. We agree that Z500 reflects not only atmospheric circulation but also the thermal structure of the atmosphere through the hydrostatic relationship. Therefore, Z500-based circulation diagnostics cannot be interpreted as purely dynamical measures.

We would like to clarify that our EOF analysis is not performed on zonal-mean Z500 itself. Instead, we first calculated Z500-based circulation activity at each latitude–longitude grid point and then took the zonal mean of the resulting activity field. Thus, zonally asymmetric circulation variability contributes to the activity measure before zonal averaging, although regional differences and opposing changes across longitude may be obscured in the zonal mean. We acknowledge that broad temperature-related variations may also influence the resulting activity patterns.

Accordingly, our EOF–regression framework identifies precipitation variability statistically associated with leading modes of Z500-based circulation activity, rather than providing a strict separation of precipitation changes into purely dynamical and thermodynamic contributions. We have clarified this interpretation and its limitations in the revised Discussion and Conclusions section.

L479–481: "Because Z500 reflects both circulation variability and the atmospheric thermal structure through the hydrostatic relationship, the Z500-based activity modes should not be interpreted as purely dynamical indicators."

Minor Comment 10

> *Figure 4: What does using GPCP or other observation-based precipitation datasets look like compared to ERA5 which is model generated?*

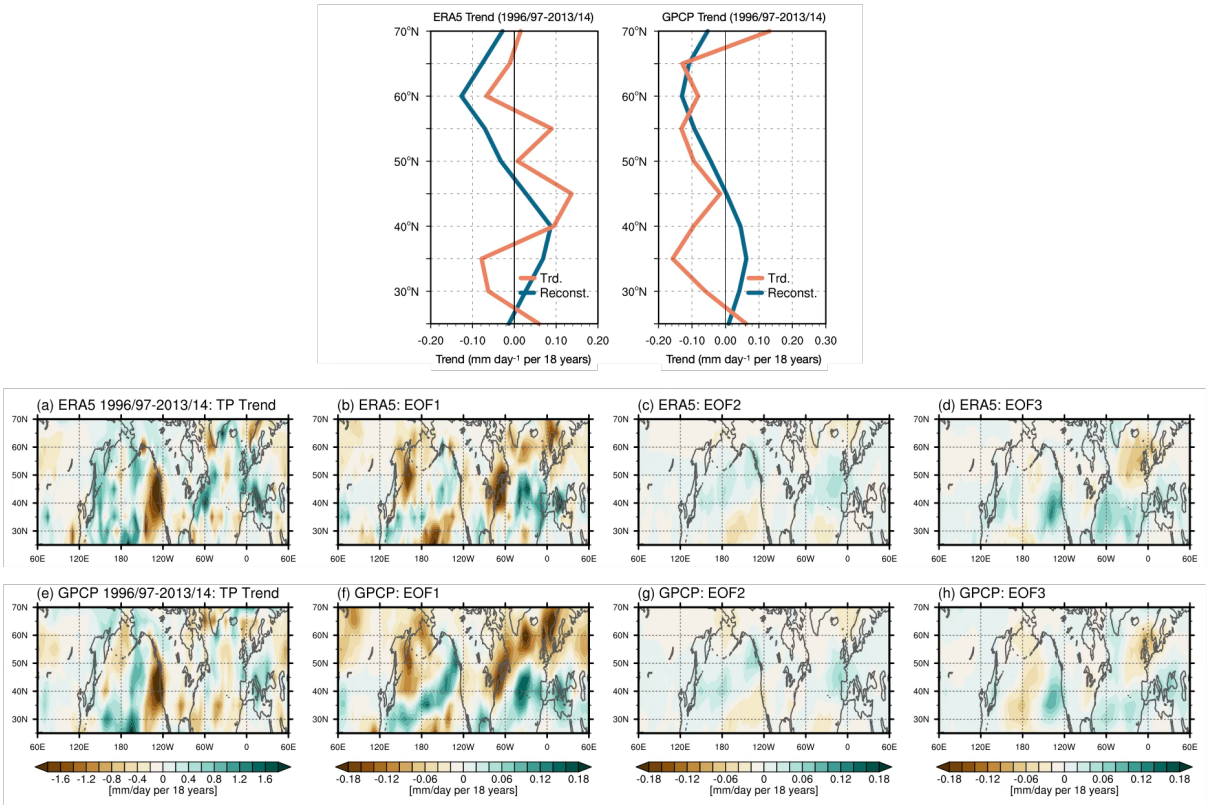
Response:

We thank the reviewer for this valuable suggestion. To assess the sensitivity of our results to the choice of precipitation dataset, we performed an additional analysis using precipitation from the GPCP.

For a consistent comparison, both ERA5 and GPCP were analyzed over the common period of 1996/97–2013/14 (18 boreal winters). Using the same ERA5-based circulation-activity EOF framework, we compared the total precipitation trends and the reconstructed precipitation trends associated with the leading three EOF modes.

The total precipitation trends exhibit notable differences between ERA5 and GPCP, particularly in their zonal-mean distributions, with a latitude-weighted PCC of 0.328. Nevertheless, the spatial distributions show some common regional features over the North Pacific and North Atlantic, although differences in magnitude and spatial extent remain. In contrast, the reconstructed precipitation trends exhibit substantially greater consistency between the two datasets. The zonal-mean reconstructions show a PCC of 0.944 and an amplitude ratio (AMP) of 1.033, indicating close agreement in both their meridional patterns and overall amplitudes. The spatial distributions of the EOF-based contributions also show broadly similar regional structures, particularly for EOF1, whereas the weaker contributions associated with EOF2 and EOF3 exhibit some regional differences.

These additional results have been included in the revised Supplementary Material as Fig. S6. This comparison provides a sensitivity assessment of the EOF-based precipitation reconstruction to the choice of precipitation dataset.



Supplementary Figure 6. Comparison of precipitation trends and EOF-based reconstructions between ERA5 and GPCP over the common period of 1996/97–2013/14 (18 boreal winters). Upper panels show the zonal-mean total precipitation trends (orange) and reconstructed precipitation trends (blue) for ERA5 (left) and GPCP (right). Lower panels show the spatial distributions of total precipitation trends and reconstructed contributions associated with EOF1–EOF3 for (a–d) ERA5 and (e–h) GPCP, respectively. The same ERA5-based circulation-activity EOF framework is used for both precipitation datasets. The latitude-weighted pattern correlation coefficients between ERA5 and GPCP are 0.328 for the total precipitation trends and 0.944 for the reconstructed trends, with corresponding amplitude ratios of 1.311 and 1.033, respectively. Units are mm day^{-1} per 18 years.

Minor Comment 11

Line 297: A single model large ensemble should probably be examined to make this claim, otherwise it is somewhat too speculative to disentangle these types of uncertainties.

Response:

In agreement with the reviewer, we have removed the statement attributing the intermodel spread to a specific source of uncertainty and revised the relevant text to avoid an overly speculative interpretation.

L467–469: "However, these mechanisms represent plausible physical explanations rather than quantitatively isolated sources of intermodel uncertainty, as the present analysis does not distinguish the contributions of internal climate variability and model structural differences."

We have also explicitly acknowledged this limitation in the revised Discussion and Conclusions section:

L489–492: "An additional limitation is that the present multi-model analysis cannot robustly separate internal climate variability from model structural uncertainty, particularly given the limited ensemble sampling. Future studies using single-model large ensembles would help quantify the contribution of

internal variability, while comparisons across multiple large ensembles would provide a more comprehensive assessment of model structural uncertainty."

Minor Comment 12

Line 360: Can you be more specific about what aspects are broadly similar between historical and SSP585 in Figure 8?

Response:

We thank the reviewer for this comment. As noted in our response to Major Comment 3, we agree that our original statement that the historical and SSP5-8.5 temperature-difference patterns in Figure 8 were "broadly similar" was not sufficiently supported. Please see our response to Major Comment 3 for a detailed description of the revised interpretation of Figure 8.

References

Blackport, Russell, and John C. Fyfe. "Climate models fail to capture strengthening wintertime North Atlantic jet and impacts on Europe." Science Advances 8.45 (2022): eabn3112.

North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J. (1982). Sampling errors in the estimation of empirical orthogonal functions. Monthly Weather Review, 110(7), 699-706.

Wills, Robert C., Michael P. Byrne, and Tapio Schneider. "Thermodynamic and dynamic controls on changes in the zonally anomalous hydrological cycle." Geophysical Research Letters 43.9 (2016): 4640-4649.

Wills, R. C., & Schneider, T. (2016). How stationary eddies shape changes in the hydrological cycle: Zonally asymmetric experiments in an idealized GCM. Journal of Climate, 29(9), 3161-3179.