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 1]

Major Comments

Major Comment 1

> *I am not convinced by the formulation used to decompose circulation variability in this study. First, while the use of either a 2-8-day band-pass filter or a 5-day running mean is totally fine for extracting transient variability, the rationale for using the former in Eq. (1) and the latter in Eq. (2) is unclear. More importantly, Eq. (3) imply that the double-prime component is defined as deviations from a 5-day running mean applied to an already band-pass-filtered field. I therefore find it difficult to understand what physical variability this component is intended to represent, while my understanding may be incorrect. Unless the physical meaning of each component is clarified and justified, I am not convinced that the resulting decomposition is physically interpretable (as stated in L370, for instance).*

Response:

We thank the reviewer for identifying this important issue. The comment prompted us to re-examine the calculation, revealing an error in the background-mean removal used in the original analysis. We had subtracted a five-day running mean, whereas the intended procedure, following Yoo et al. (2025), uses a 31-day running mean. We have corrected this step and repeated the affected analyses, updating Figures 1–8 and Supplementary Figures 1–4, along with the corresponding numerical values in the text. This includes re-evaluating the PROJ-based model subsets used in Figures 6–8. We have also entirely revised Section 2.2 to correct the equations, distinguish the processing steps, and clarify the physical meaning of each term. Although the quantitative values have changed, the principal findings and conclusions remain unchanged, as explained below. We apologize for the error and the lack of clarity in the original description.

The corrected procedure follows the daily adaptation by Yoo et al. (2025) of the monthly storm-track diagnostic used by Lau (1988). Lau calculated monthly root-mean-square (RMS) statistics of band-pass-filtered geopotential-height fluctuations relative to their monthly means. Yoo et al. adapted this approach to daily sampling using a five-day RMS window and a 31-day running mean as the background reference. We follow their background-removal and averaging procedure, while retaining the squared activity measure to permit an exact additive decomposition.

The three operations in our analysis serve distinct purposes. First, the 2–8-day band-pass filter extracts synoptic transient disturbances from the geopotential-height anomalies, with the remaining field defined as the residual component. Second, subtraction of the 31-day running mean defines departures from the evolving monthly-scale background. Third, averaging the squared departures over five days provides a daily measure of their local activity. The five-day averaging window and the 31-day background window are centered on the same evaluation day, and the background reference is held fixed across the

five samples. Thus, the five-day window determines the averaging interval for activity, rather than the background from which the fluctuations are defined.

For the band-pass-filtered component, the 31-day background subtraction provides a consistent local reference, since monthly and longer-timescale variability is already strongly suppressed by the filter. For the total and residual fields, it removes the monthly-scale background and measures departures from it. Applying the same linear background-removal operation to all components preserves their additive relationship. Squaring and averaging therefore yield an exact decomposition of total activity into the transient contribution, the residual contribution, and their cross term.

The revised Section 2.2 explicitly defines these quantities and clarifies their interpretation. The transient contribution, VAR_T, measures storm-track activity through the mean-square amplitude of synoptic disturbances. The residual contribution, VAR_S, measures the mean-square departure of the residual circulation, primarily slowly varying disturbances, from its 31-day background. We now refer to this component as the "low-frequency component" rather than the "stationary component" used in the original manuscript. The cross term, $2\mathrm{COV}$, quantifies whether the two components reinforce or partially cancel one another in the total activity.

The corrected analyses preserve the principal findings. The transient component remains the dominant contributor to the precipitation reconstruction, and its contributions associated with shifts and sharpening remain important in ERA5. In the multi-model ensemble, differences in modal contributions across models continue to weaken the ensemble-mean reconstruction through cancellation. The smaller relative magnitude of the circulation-associated reconstruction under SSP5-8.5 also persists. These findings are therefore supported by the corrected calculation using the monthly-scale background reference. We appreciate the reviewer's comment, which led us to correct the analysis and substantially improve the formulation and physical interpretation presented in Section 2.2.

Major Comment 2

> *Definition of anomalies: This study defines anomalies as deviations from DJF-mean climatology. This means that anomalies include seasonal march embedded in climatological-mean fields as well as high-frequency and low-frequency components of deviations from climatological-mean fields on each calendar day. It would be fine if the authors intend to include the seasonal march of climatology, but in this case, the trends should be interpreted as the combinations of the contributions of high-frequency (sub-weekly) and low-frequency variabilities of anomalies, and low-frequency seasonal march in the climatological-mean fields, and their interactions.*

Response:

In the original manuscript, the anomalies defined relative to a fixed DJF climatology retained the climatological seasonal march. We agree that this complicates the interpretation of the activity measures, because, as the reviewer wrote, they can include contributions from the seasonal cycle and its cross terms with departures from the calendar-day climatology.

We have repeated the analysis after subtracting the calendar-day climatology from the geopotential-height fields. This change was implemented together with the correction to the 31-day background-mean removal described in our response to Major Comment 1. The two operations serve distinct purposes: calendar-day climatology removal excludes the climatological seasonal march, whereas subtraction of the 31-day running mean defines departures from the evolving background within each

individual winter. The subsequent five-day average of squared departures provides the daily activity measure.

The revised calculation produces smaller reconstructed precipitation-trend amplitudes than the original analysis. Nevertheless, the ERA5 reconstruction retains the broad meridional contrast between subtropical drying and midlatitude wetting (Fig. 1). Pattern agreement remains strong, with PCC values of 0.90, 0.75, and 0.82 for ERA5, the historical multi-model mean, and the SSP5-8.5 multi-model mean, respectively. However, the corresponding amplitude ratios (AMP) are 0.43, 0.27, and 0.21. These results indicate that the retained circulation-activity modes capture important features of the meridional precipitation-trend structure while representing a more limited portion of its magnitude than previously reported. The updated modal decomposition also shows that the shift-related mode, EOF2, provides the principal transient-associated contribution in ERA5, whereas the sharpening-related contribution is weak (Figs. 4–5). We have therefore revised both the amplitude interpretation and the description of the relative modal contributions, rather than retaining the original conclusions unchanged.

We have revised Section 2.2 to define anomalies relative to the calendar-day climatology. We have also updated the affected figures and revised the abstract, results, and conclusions to reflect the reduced reconstructed amplitude and updated modal contributions. In particular, we have removed statements implying close amplitude agreement or near-unity reconstruction amplitude and now distinguish pattern agreement from amplitude recovery. The principal changes are listed below.

L16–19: "In ERA5, the leading circulation modes capture the broad meridional contrast between subtropical drying and midlatitude wetting, but underestimate the precipitation-trend amplitude. Among the transient-activity modes, the shift-related mode provides the largest contribution to the reconstructed trend, while the pulsing and sharpening contributions are weaker."

L103–105: "Following the storm-track diagnostic approach of Lau (1988), we quantify circulation variability using the mean-square activity of 500-hPa geopotential-height anomalies (Z'_{500}). The anomalies are defined by removing the calendar-day climatological mean field, calculated over the 1979/80–2013/14 reference period, at each grid point."

L199–203: "We first assess how well the leading circulation-variability modes reconstruct multidecadal midlatitude precipitation trends. Applying the EOF–regression framework to ERA5 captures the broad DJF precipitation-trend pattern but underestimates its amplitude (Fig. 1a). Using nine circulation modes (three EOFs each for $\mathrm{VAR_S}$, $\mathrm{VAR_T}$, and COV), the reconstruction (blue line) agrees well with the total precipitation trend (red line) in spatial structure (PCC = 0.90) but recovers less than half of its amplitude (AMP = 0.43; NRMSE = 0.64)."

L210–212: "Structural agreement is reasonably strong (PCC = 0.75), indicating that the CMIP6 MMM broadly captures the dynamical linkage between circulation change and hemispheric precipitation redistribution. The reconstructed amplitude is, however, only a small fraction of the simulated trend (AMP = 0.27; NRMSE = 0.91)."

L220–224: "Here, dynamical consistency is defined as the degree to which the reconstructed circulation-driven precipitation trend matches the total simulated precipitation trend, as quantified by the NRMSE score (Eq. (9)). For the four models with the lowest NRMSE scores (Fig. 1d), the reconstructed (blue dashed line) and simulated trends (orange dashed line) agree much more closely in both pattern (PCC = 0.90) and magnitude (AMP = 0.62; NRMSE = 0.60), and the intermodel spread is substantially reduced (shading)."

L229–233: "In SSP5-8.5 (Fig. 1c), the reconstruction (blue line) still captures the dominant spatial structure of the total precipitation trend (orange line). The SSP5-8.5 MMM exhibits reasonable structural agreement (PCC = 0.75), while its amplitude remains small (AMP = 0.27; NRMSE = 0.91).

Restricting the SSP5-8.5 MMM to the four lowest-NRMSE models improves the agreement (PCC = 0.86, AMP = 0.34, and NRMSE = 0.74; Fig. 1e and Supplementary Fig. 2e), indicating that intermodel differences still contribute to the reduced amplitude in the full MMM."

L248–250: "ERA5 (orange asterisk in Fig. 2a) exhibits high pattern agreement but an amplitude ratio substantially below unity, indicating that the reconstruction captures the broad spatial structure more successfully than its magnitude."

L298–301: "In ERA5, EOF1 (panel b), representing storm-track pulsing, produces relatively weak regional contributions. EOF2 (panel c), associated with meridional shifting, provides the largest transient-associated contribution and exhibits a pronounced contrast between subtropical drying and midlatitude wetting. The EOF3 contribution (panel d) is weak. Thus, the ERA5 reconstruction is dominated by the shift-related mode."

L429–433: "The retained circulation-activity modes capture important features of the meridional precipitation-trend pattern, while recovering only part of its amplitude. In ERA5, the transient-associated reconstruction is dominated by the shift-related mode, with a much weaker sharpening contribution. These results support the framework as a compact diagnostic of circulation-associated precipitation redistribution, while limiting claims about the magnitude of precipitation change represented by the retained modes."

Major Comment 3

> *The authors use different numbers of models for historical and ssp585 simulations. Given the presumably larger inter-model spread in long-term trends than in climatological-means, the difference in the cohorts of models may lead to artificial signals. I recommend that the authors should use the same cohort of models or, show the results based on the same cohort of models in the Supplement.*

Response:

We thank the reviewer for this important suggestion. We agree that differences in the model ensembles used for the historical and SSP5-8.5 simulations may introduce sampling-related effects, particularly given the substantial intermodel spread in long-term precipitation trends. To evaluate this sensitivity, we repeated the analysis using the same 12 CMIP6 models available for both periods: 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. Each model was assigned equal weight in calculating the multi-model mean.

The common-model analysis yields results broadly consistent with those obtained from the full ensembles (21 historical and 19 SSP5-8.5 models). As a representative example, Figure R1 compares the PCC and AMP of the circulation-based precipitation reconstructions using the common 12-model ensemble. The fitted PCC–AMP relationships retain positive slopes in both the historical and SSP5-8.5 simulations, consistent with the full-ensemble results (Fig. 2). Specifically, the regression slope decreases from 0.61 to 0.58 (approximately 3.8%) for the historical simulations and from 2.95 to 2.73 (approximately 7.5%) for SSP5-8.5. The substantially steeper positive relationship in SSP5-8.5 is therefore maintained when the same models are used for both periods.

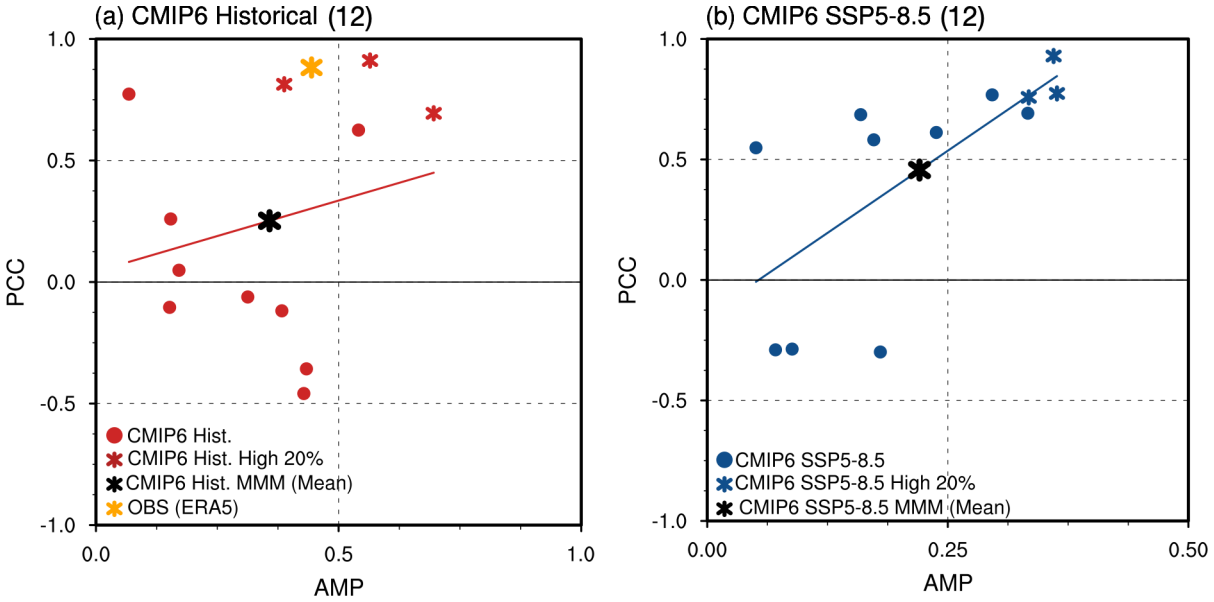


Figure R1. PCC–AMP relationship for circulation-based precipitation reconstructions using the 12 CMIP6 models common to the historical and SSP5-8.5 simulations. Results are shown for (a) historical and (b) SSP5-8.5 simulations. Colored circles indicate individual CMIP6 models, colored asterisks indicate 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 arithmetic means of individual-model PCC and AMP values also show similar characteristics. For the historical simulations, the mean PCC and AMP change from 0.32 and 0.39 in the full ensemble to 0.25 and 0.36 in the common-model ensemble, respectively. For SSP5-8.5, the corresponding values remain nearly unchanged (PCC: 0.45 to 0.46; AMP: 0.22 to 0.22). Thus, the contrast between the historical and SSP5-8.5 simulations, characterized by higher mean PCC but lower mean AMP in SSP5-8.5, persists in the common-model analysis.

These results indicate that the principal characteristics of the PCC–AMP relationship are not substantially affected by differences in model composition between the two experiments. Although individual model statistics vary with the selected ensemble, the positive relationships, their relative slopes, and the contrasting behavior of pattern agreement and amplitude recovery remain consistent. We present the common-model sensitivity results here to directly address the reviewer's concern regarding model sampling.

Major Comment 4

> *It would be much more meaningful to apply EOF analysis to 2D fields instead of zonal mean fields. Related to this, readers cannot understand the patterns of leading EOF modes of circulation variabilities, as there are no figures showing these. Please add these figures.*

Response:

Our choice of zonal-mean EOFs reflects our aim to develop a compact framework for interpreting Northern Hemisphere meridional precipitation redistribution. The leading transient-activity modes summarize variations in storm-track amplitude, latitude, and meridional concentration, which we interpret as pulsing, shifting, and sharpening. This representation facilitates comparison of their

contributions across reanalysis, models, and simulation periods. Latitude–longitude EOFs provide complementary information by retaining regional differences in storm-track position, intensity, and geometry. Such analyses can be performed over either hemispheric or basin domains (Wettstein and Wallace, 2010). However, the additional spatial structure can distribute variability among regional modes that are largely confined to individual basins (Fig. R2; in ERA5, for example, EOF1 is concentrated over the North Atlantic and EOF3 over the North Pacific). The ordering and interpretation of such modes may not be consistent across datasets, which would complicate a mode-by-mode comparison between reanalysis and models. We therefore retain the zonal-mean framework for the main analysis because it offers a compact and readily interpretable description at the hemispheric scale.

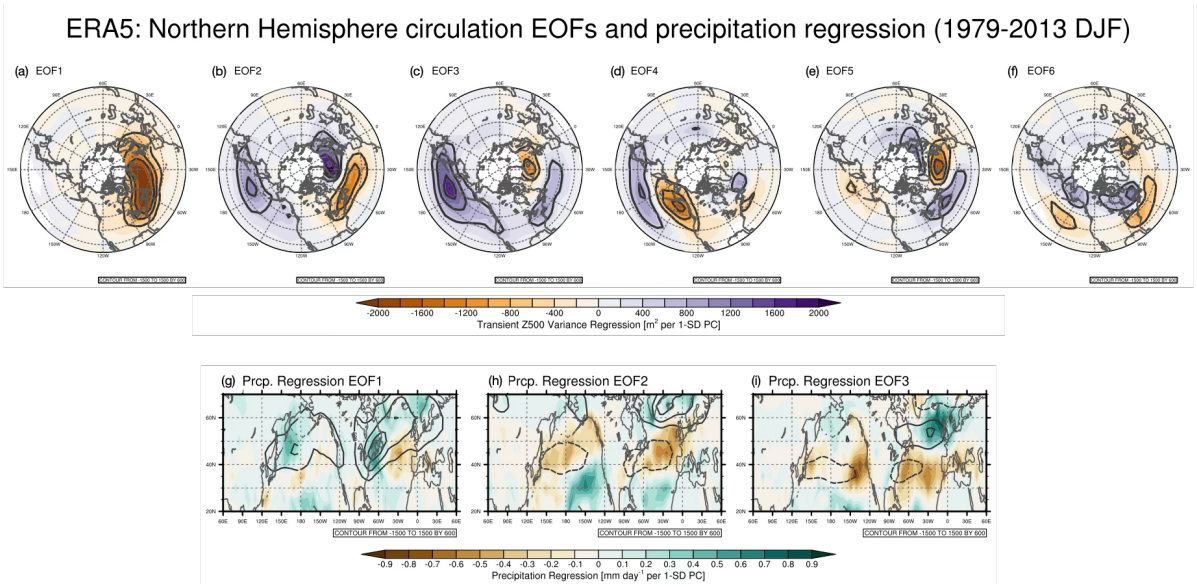


Figure R2. Spatial regression pattern of Northern Hemisphere circulation activity and precipitation associated with the leading two-dimensional EOF modes in ERA5. (a–f) Latitude–longitude distributions of transient Z500 variance regressed onto the standard principal components (PCs) of EOF1–EOF6, respectively. (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.

Nonetheless, we agree with the reviewer that the spatial patterns associated with the leading EOFs help readers understand the regional expression of each mode. We have therefore 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. The regression patterns associated with EOF1 show pulsing of storm-track activity in both the North Pacific and North Atlantic, while those associated with EOF2 and EOF3 show a poleward shift and meridional sharpening, respectively. Precipitation anomalies are consistent with the associated storm-track anomalies. Because these maps are obtained by regression at each latitude–longitude grid point onto PCs derived from the zonal-mean EOF analysis, they depict the regional expression of each zonal-mean mode. The coherence between basins is a feature of the resulting regression patterns, rather than a constraint imposed by the regression. As a complementary analysis, we also performed EOF analysis directly on two-dimensional latitude–longitude fields of transient Z500 variance and present the associated spatial regression patterns of circulation activity and precipitation in Fig. R2. Such EOFs can additionally identify regional modes that are weakly represented in the zonal mean, including opposing changes between basins.

The final paragraph of the original manuscript briefly acknowledged the scope of the zonal-mean framework and identified basin-scale analysis as a complementary extension. We agree that this distinction should be made clearer when introducing the method and interpreting the results, and have therefore revised Sections 2.3 and 3.2 as follows:

L129–135: “EOF analysis is applied independently to the zonal-mean activity fields of each component, so that the resulting EOFs depend only on latitude. This approach provides a compact description of the meridional structure of Northern Hemisphere circulation variability and facilitates comparison of the associated precipitation responses across the reanalysis, models, and simulation periods. The retained PCs summarize hemispheric activity variations but do not describe all independent regional variability. Although the analysis could be extended to retain longitudinal structure, doing so would require a basin-scale treatment and the attribution of regional-scale variability and trends, which are beyond the scope of this study.”

L302–305: “The longitude–latitude structures of storm-track activity and precipitation regressed onto the corresponding PCs (Supplementary Fig. 5) also reveal consistent pulsing, shifting, and sharpening of storm-track activity in both the North Pacific and North Atlantic, accompanied by coherent precipitation anomalies.”

Major Comment 5

> *The use of PROJ (Eq. 8) as a combined measure of the similarity and amplitude of precipitation trends requires the reconstructed trends to always underestimate the observed trends. The authors might want to use another metric that works regardless of the amplitude relative to the observations.*

Response:

We thank the reviewer for identifying this limitation of the PROJ metric. We agree that PROJ, defined as the product of pattern correlation and amplitude ratio, is not a direct measure of reconstruction error. We therefore introduced the normalized root-mean-square error (NRMSE) as the primary metric for evaluating the reconstructed precipitation trends:

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{\phi}\left([\Delta P'_{\mathrm{Reconst.}}](\phi) - [\Delta P'](\phi)\right)^2 \cos\phi}{\sum_{\phi}[\Delta P']^2\,(\phi)\cos\phi}}.$$

When PCC and AMP are defined using the same weighted inner product, this can be written as

$$\mathrm{NRMSE} = \sqrt{1 + \mathrm{AMP}^2 - 2\,\mathrm{PCC}\cdot\mathrm{AMP}} = \sqrt{1 + \mathrm{AMP}^2 - 2\,\mathrm{PROJ}}$$

NRMSE directly measures the normalized distance between the reconstructed and target trend fields. A value of zero indicates perfect reconstruction, and smaller values indicate better reconstruction. Unlike PROJ, NRMSE penalizes both underestimation and overestimation of the trend magnitude.

We reevaluated the models using NRMSE for the total circulation-based precipitation reconstruction and for the reconstructions associated with the individual EOF modes in both the historical and SSP5-8.5 experiments. Overall, the rankings based on NRMSE were broadly consistent with those based on PROJ, particularly for the SSP5-8.5 simulations and the individual EOF reconstructions. However, some changes in model rankings were identified. GFDL-CM4 (AMP = 1.37) fell from third place under PROJ to fourteenth under NRMSE. Similarly, MPI-ESM1-2-HR (AMP = 1.35) fell from seventh to nineteenth. These examples confirm the reviewer’s concern that PROJ can favor models that reproduce the spatial pattern but overestimate its amplitude. Nonetheless, the overall model-evaluation results and the principal conclusions remain robust to the choice of metric. We therefore use NRMSE as the primary metric for ranking the models in the revised manuscript. PCC and AMP are retained as complementary

diagnostic metrics because they distinguish pattern similarity from amplitude error. PROJ is retained only for comparison with the original analysis.

We revised the Methods section to define NRMSE and explain its relationship to PCC, AMP, and PROJ. We also updated the model-ranking tables and corresponding figures using NRMSE and revised the Results section to discuss the differences between pattern and amplitude errors.

Other Comments

Other Comment 1

> *L56: I am afraid that I could not grasp the conceptual meaning of "circulation reorganization". Please clarify its meaning and relevance to the present study.*

Response:

We intended this term to describe changes in the intensity and spatial distribution of circulation variability, particularly changes in storm-track amplitude, meridional position, and meridional width. We have clarified the phrase as follows:

L58–59: "…, CMIP5 and CMIP6 exhibit substantial spread in changes in storm-track intensity, meridional position, and width …"

Other Comment 2

> *Please add the units of the trends to the figures (e.g., Figs. 1, S1, and S4).*

Response:

We thank the reviewer for this suggestion. We have reviewed the physical units in all figures and added the corresponding trend units to all figures presenting trends, including those in the main manuscript and Supplementary Material. We have also ensured that the units are consistently reported in the figure labels, color bars, and captions.