Replies to Reviewer #1:

The manuscript "Low-Frequency Stratospheric Variability and its Link to Surface Climate: Insights from Reanalysis and Multi-Century Coupled Climate Model Simulations" by Wang et al. analyses long-term variability of the wintertime stratosphere in the Northern Hemisphere using the CERA-20C reanalysis and multi-century simulations from three versions of the CESM model, including the high-top WACCM version and two low top versions CAM4 and CAM6. The authors identify a 15-years long oscillation in the stratospheric NAM index and find signatures of this oscillation in sea level pressure in both the reanalysis and the model simulations. Spectral and wavelet analyses also identify a 30-years long oscillation in the models; however, the authors recommend interpreting this result with caution.

While the authors present a solid statistical analysis and the finding of a clear peak around 15 years in the NAM spectra (Fig. 1) is potentially interesting, I am not convinced that the manuscript, in its current form, makes a sufficiently useful contribution to the field. I suggest that the authors elaborate further on their findings and substantially revise the manuscript. Below is a list of issues that I believe should be addressed before possible publication.

Response: We thank the reviewer for taking the time to review our manuscript and to provide insightful comments. Please see our detailed point-by-point response below.

1. What is the mechanism for multi-year stratospheric NAM variability? I tend to think about the stratospheric variability that every autumn the extratropical stratosphere starts from a clean state, a new polar vortex is formed and no memory from the previous stratospheric winter is carried over. Any multi-year extratropical stratospheric variability is externally generated – either from the tropical stratosphere (QBO) or from the troposphere via Rossby wave propagation. Scott and Haynes (1998) proposed a mechanism involving low-latitude winds as a medium for interannual stratospheric memory, but I do not think this mechanism is suitable to explain a 15-years long oscillation. And if the source of the oscillation is outside the polar stratosphere, then what is the role of the polar stratosphere? Although the authors state in the abstract that the 15-years long stratospheric variability "tends to lead NAO surface variability"

I would hesitate to make such a statement until the mechanism is understood. Therefore, I suggest the authors propose a mechanism that can explain their results.

*Reference: Scott, R. K., and P. H. Haynes, 1998: Internal interannual variability of the extratropical stratospheric circulation: The low-latitude flywheel. Quart. J. Roy. Meteor. Soc., **124**, 2149–2173, https://doi.org/10.1002/qj.49712455016*

Response: We thank the reviewer for these suggestions. We agree that the original statement that the stratospheric NAM "*tends to lead NAO surface variability*" implies a direction of influence that cannot be established from our lead–lag analysis alone. We have therefore revised the Abstract, Main text, and Conclusions to remove wording implying that stratospheric variability leads or drives the NAO and instead describe only their low-frequency statistical association (**please see Line 9, 220, 365**).

We also agree with the reviewer that understanding the physical origin of the ~15-year stratospheric variability is important. Following the reviewer's suggestion, we conducted an exploratory analysis to examine whether the QBO might be a source of this variability. Motivated by Dimdore-Miles et al. (2021), who found that long-term modulation of the QBO amplitude is associated with multidecadal variability in sudden stratospheric warming occurrence on ~60–90-year timescales, we also examined whether the QBO amplitude is a possible source of stratospheric ~15-year variability. We calculated the QBO amplitude from equatorial (5°S–5°N) zonal-mean zonal winds. After removing the monthly climatology and linear trend, the monthly winds were band-pass filtered over 18–40 months, and the amplitude envelope was obtained using the Hilbert transform. The resulting amplitude was averaged over September–November (SON) and smoothed using the same Gaussian filter as the NAM index.

Figure R1 (new Figure A7) shows the cross-wavelet power and coherence between autumn 30 hPa QBO amplitude and winter stratospheric NAM in ERA5 (1940-2025) and CERA-20C. Cross-wavelet power identifies periods and times when the two indices exhibit enhanced common spectral power, whereas wavelet coherence measures the strength of their relationship in time–frequency space (Grinsted et al., 2004; Dimdore-miles et al., 2021). The cross-wavelet power shows some enhancement near the ~15-year period in ERA5 and CERA-20C, although the signal

does not reach statistical significance in either reanalysis (Figure R1c, d). The wavelet coherence also does not show a statistically significant relationship at this timescale (Figure R1e, f). We also repeated the analysis using QBO amplitude indices at 15, 20 and 50 hPa and obtained similar results (not shown). Although these results do not show a statistically significant relationship between QBO amplitude and the ~15-year stratospheric variability, they do not rule out a possible contribution from the QBO. The signal could be too weak to reach statistical significance given the limited number of years available, it could be intermittent, or the relationship may not exist. We have added Figure R1 as new Figure A7 and included a discussion of these results in the Discussion Section (**please see Lines 422-428**).

We acknowledge that the physical origin of the ~15-year stratospheric variability and the role of the polar stratosphere remain unclear. The primary focus of this study is to document the ~15-year stratospheric variability identified in CERA-20C and three climate model simulations, and to examine its statistical association with surface circulation. Identifying the underlying physical mechanism and clarifying the role of the polar stratosphere will require further dynamical analyses and targeted model experiments, which we consider an important direction for future research.

We have therefore added a paragraph to the Discussion addressing possible sources of the ~15-year variability and the limitations of our analysis (**please see the final 3 paragraphs in Discussions, Line 416-440**): "*First, the physical origin of the ~15-year stratospheric variability is not yet established. Possible sources include low-frequency variations in tropospheric planetary wave forcing, potentially influenced by oceanic variability, and influences from the tropical stratosphere, such as a modulation of the quasi-biennial oscillation (QBO). Further internal stratospheric processes may also contribute. For example, Scott and Haynes (1998) proposed that anomalies in low-latitude stratospheric winds can persist between winters and influence the polar vortex in subsequent winters. However, whether this mechanism could explain variability on ~15-year timescales remains unclear…… Our study documents the ~15-year stratospheric variability across reanalysis data and climate model simulations and characterizes its statistical association with surface circulation. Further dynamical analyses and targeted model experiments are needed to address these unresolved questions.*"

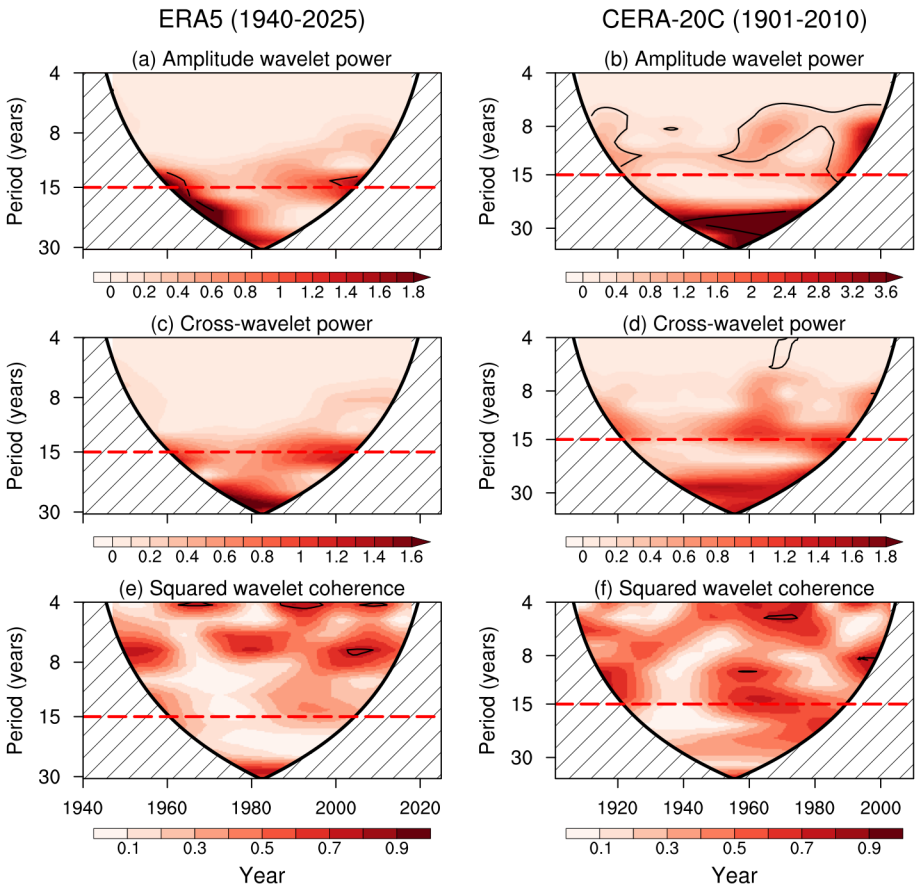


Figure R1. Wavelet analysis of the 30-hPa QBO amplitude and stratospheric NAM in ERA5 (left; 1940–2025) and CERA-20C (right; 1901–2010). (a, b) Wavelet power of the QBO amplitude, (c, d) cross-wavelet power between QBO amplitude and NAM, and (e, f) squared wavelet coherence. The QBO amplitude is averaged over September–November (SON) and NAM over November–March (NDJFM). Both indices are standardized and smoothed. Thin black contours indicate 95% pointwise significance. Thick black curves mark the cone of influence, with hatched regions indicating areas affected by edge effects. Red dashed lines indicate the 15-year period.

Reference:

Dimdore-Miles, O., L. Gray, and S. Osprey, 2021: Origins of multi-decadal variability in sudden stratospheric warmings, Weather and Climate Dynamics, 2, 205–231, https://doi.org/10.5194/wcd-2-205-2021.

Grinsted, A., J. C. Moore, and S. Jevrejeva, 2004: Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlin. Processes Geophys., 11, 561–566, https://doi.org/10.5194/npg-11-561-2004.

Scott, R. K., and P. H. Haynes, 1998: Internal interannual variability of the extratropical stratospheric circulation: The low-latitude flywheel. Quart. J. Roy. Meteor. Soc., ***124****, 2149–2173, https://doi.org/10.1002/qj.49712455016*

2. I am concerned by the lack of agreement between the surface responses in the reanalysis and in the models at lag 0 (Figure 5), and I believe the authors should be concerned too. While the Atlantic responses, which essentially indicates that strong/weak polar vortex corresponds to positive/negative NAO, are in a reasonable agreement between the reanalysis and the models, the Pacific response in Figure 5 is quite different. Both model versions show an anomalous high in the northern Pacific. The high corresponds to a weakened Aleutian low, and, likely, a weakened planetary wave driving, which would be consistent with a strong polar vortex and positive NAM. On the contrary, CERA-20C shows a low in the northern Pacific. If I understand correctly, the lag 0 means a correlation between simultaneous seasonal (NDJFM) mean NAM and seasonal mean gridded SLP. The physical coupling is represented by upward Rossby wave propagation and interaction with the stratospheric zonal mean winds (upward coupling) and tropospheric circulation adjustment to stratospheric zonal mean anomalies (downward coupling). Both these processes occur on a weekly timescale, so there are several individual events that contribute to the seasonal means. The result is a superposition of the forcing fields (upward coupling) and the response fields (downward coupling). The different signatures seen in Figure 5 suggests that the relative magnitude of the upward and downward coupling is different between the models and the reanalysis. Or how do the authors explain the different signatures? And what are the implications of these differences for the conclusions of the manuscript? I think this should be discussed in the manuscript.

Response: Thank you for this comment. The zero-lag North Pacific SLP patterns differ between the reanalysis and climate models (Figure 5), but the reason for this discrepancy remains unclear. In CERA-20C, a stronger polar vortex is associated with a deepened Aleutian Low at zero lag, opposite to the commonly reported relationship on subseasonal-to-seasonal timescales (weaker Aleutian Low- stronger polar vortex). Some previous studies also found a stronger Aleutian Low associated with a stronger polar vortex on decadal-to-multidecadal timescales (e.g., Hu and Guan, 2018; Seabrook et al., 2023; Palmeiro et al., 2025). However, these findings do not explain why the NAM-Aleutian Low relationship differs between the reanalysis and

models. Differences between models and reanalysis in the North Pacific surface response have also been reported on subseasonal timescales. Dai et al. (2024) found that most CMIP6 models simulate a cyclonic North Pacific SLP response following sudden stratospheric warmings, whereas reanalysis shows no robust response in this region. Garfinkel et al. (2025) also found that many subseasonal forecast systems overestimate the near-surface response to stratospheric vortex anomalies.

As the reviewer suggested, our lead-lag regression patterns may reflect both upward tropospheric wave forcing and downward stratospheric influences, whose relative contributions may differ between the reanalysis and models. However, our monthly data cannot resolve these processes on weekly timescales or distinguish their relative contributions. We have added a discussion of this uncertain North Pacific pattern (**Line 429-438**): "*Second, the low-frequency relationship between stratospheric NAM variability and the Aleutian Low remains uncertain, given the lack of agreement between the reanalysis and climate model simulations…Our lead-lag regressions are based on winter-mean indices and may reflect the combined influence of upward tropospheric wave forcing and downward stratospheric influences, which occur on weekly timescales. One possibility is that the relative contributions of these upward and downward processes differ between the reanalysis and models. However, our monthly data cannot resolve these coupling processes on weekly timescales or distinguish their relative contributions.*"

This discrepancy does not change the main conclusions of our manuscript. In the original Abstract, we already stated that (**please see Lines 13-14**) "*the relationship between the stratospheric ~15-year variability and the Aleutian Low is comparatively weak and less consistent across the reanalysis and models.*" We also noted in the original Discussion that (**please see Lines 367-370**) "*By contrast, the corresponding relationships with North Pacific variability are comparatively weak and less consistent between the reanalysis and model simulations, suggesting that the Pacific linkage to the stratospheric ~15-year variability period is more uncertain than the Atlantic linkage*" (Figures 4 and 5). We have also revised the paragraphs discussing the relationship between the stratospheric ~15-year variability and North Pacific surface circulation to more explicitly address this discrepancy (**please see Lines 254-279**).

We also repeated the lead–lag regressions between stratospheric NAM and tropospheric Aleutian Low using ERA5 for 1959–2025 with more extensive upper-air observational coverage (Figure R3). At zero lag, the regression coefficient in ERA5 is weak and not statistically significant (Figure R3a). The SLP regression pattern near lag 0 shows positive anomalies near the Bering Strait and negative anomalies farther south over the North Pacific (Figure R3b), differing from the more widespread negative SLP anomalies in CERA-20C. Thus, ERA5 does indeed not reproduce the deepened Aleutian Low associated with a stronger polar vortex found in CERA-20C, further highlighting the uncertainty in the low-frequency relationship between the stratospheric NAM and the North Pacific circulation.

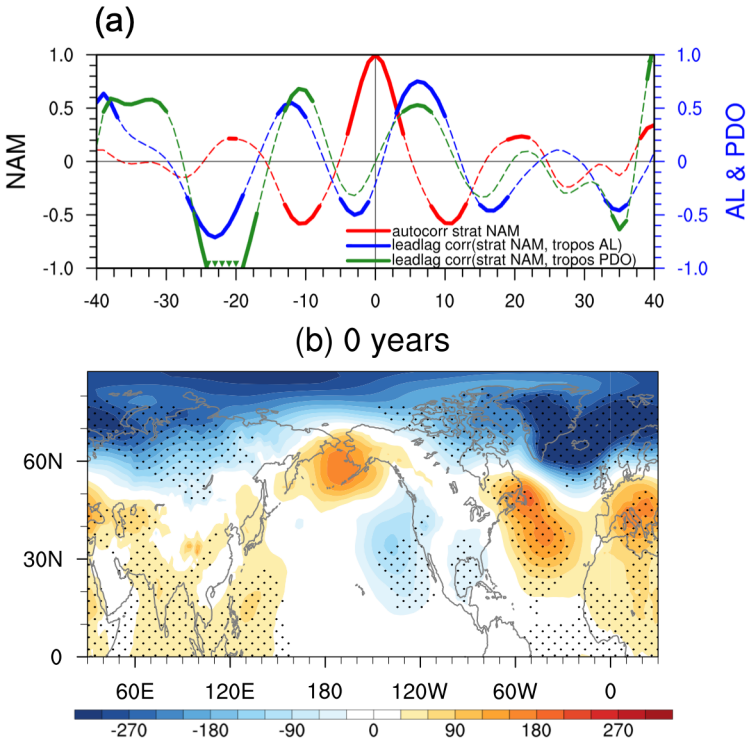


Figure R3. (a) Similar to Figure 4a, but for ERA5 (1959-2025). (b) Lagged regression patterns of SLP with the stratospheric NAM index for ERA5 at zero lag, averaged over years -2~2.

Reference:

Dai, Y., P. Hitchcock, and I. R. Simpson, 2024: What drives the spread and bias in the surface impact of sudden stratospheric warmings in CMIP6 models? J. Climate, 37, 3917–3942, https://doi.org/10.1175/JCLI-D-23-0622.1

Garfinkel, C. I., and Coauthors, 2025: A process-based evaluation of biases in extratropical stratosphere–troposphere coupling in subseasonal forecast systems. Wea. Climate Dyn., 6, 171–195, https://doi.org/10.5194/wcd-6-171-2025

Hu, D., and Z. Guan, 2018: Decadal relationship between the stratospheric Arctic vortex and Pacific decadal oscillation. J. Climate*, **31**, 3371–3386, https://doi.org/10.1175/JCLI-D-17-0266.1.*

Seabrook, M., D. M. Smith, N. J. Dunstone, R. Eade, L. Hermanson, A. A. Scaife, and S. C. Hardiman, 2023: Opposite impacts of interannual and decadal Pacific variability in the extratropics. Geophys. Res. Lett.*, **50**, e2022GL101226, https://doi.org/10.1029/2022GL101226.*

Palmeiro, F. M., P. Hitchcock, and J. García-Serrano, 2025: Separating multi-decadal and interannual dynamics in the North Pacific–stratospheric polar vortex relationship and their teleconnection to the North Atlantic. Geophys. Res. Lett.*, **52**, e2025GL116639, https://doi.org/10.1029/2025GL116639.*

3. The previous comment leads to the next question – is the stratosphere-troposphere coupling in the CERA-20C reanalysis realistic? I understand that the authors need long timeseries to quantify the multi-decadal variability, but I believe that ERA5 timeseries would be sufficient for testing the 15-years long oscillation. More importantly, I also suggest using ERA5 for testing the stratosphere-troposphere coupling at lag 0 as well as at lags +/- 15 years. This would provide a benchmark against which the models and the CERA-20C could be validated.

Response: We thank the reviewer for this suggestion. We have repeated the wavelet and lead–lag analyses using ERA5 (new Figure A5; shown here as Figure R4).

The ERA5 wavelet analysis shows enhanced power around 15–20-year timescales during the late 1980s and 1990s (Figure R4a). CERA-20C also exhibits intermittent power near 15 years, mainly from the 1940s to the 1980s. Thus, both reanalyses show intermittent low-frequency stratospheric variability, although the enhanced power occurs during different periods and at slightly longer timescales in ERA5.

We also examined the lead–lag regressions between the stratospheric NAM and the tropospheric NAO, Aleutian Low, and PDO indices (Figure R4b, c). For the ERA5 lead–lag regressions, we use 1959–2025 to focus on the period with more extensive upper-air observational coverage. In ERA5, the stratospheric NAM is positively correlated with the NAO near zero lag, consistent with the relationship found in CERA-20C. However, the NAM–NAO regression in ERA5 does not show a clear relationship at lags of ±15 years. Over the North Pacific, the NAM correlations with the Aleutian Low and PDO indices are relatively weak and not statistically significant at zero lag, which is not consistent with CERA-20C. These results further support that the low-frequency relationships between the stratospheric NAM and North Pacific circulation are less consistent across reanalysis and models.

Overall, the ERA5 results support the positive NAM–NAO relationship near zero lag found in CERA-20C, but the NAM–NAO relationships at lags of ±15 years differ between the two reanalyses. As CERA-20C assimilates surface pressure and marine wind observations in the atmosphere but does not assimilate upper-air observations, we agree that we should discuss the CERA-20C results compared with ERA5 in the manuscript. We have included these additional analyses in the Supplementary Material (new Figure A5) and added the following discussion (**Lines 402-415**): "*Although the long CERA-20C record is useful for examining stratospheric multi-year to multi-decadal variability, the results should be interpreted with caution. CERA-20C assimilates surface pressure and marine wind observations but does not assimilate upper-air observations (Laloyaux et al., 2018). Thus, the stratospheric variability in CERA-20C is not directly constrained by upper-air observations. For comparison, we therefore also examine ERA5 from 1959-2025, which assimilates upper-air observations (Hersbach et al., 2020), as well as NOAA 20CR from 1837-2015, which assimilates surface pressure observations but no upper-air observations (Slivinski et al., 2019). ERA5 and NOAA 20CR show intermittent spectral power at periods of approximately 15–20 years (Figure A5a, d) and positive NAM–NAO regression coefficients near zero lag (Figure A5b, e). In addition, the zero-lag SLP pattern in ERA5 shows a positive NAO-like dipole (Figure A6b), consistent with that in CERA-20C (Figure 5a2). However, the positive NAM–NAO regression that was found near +15 years in CERA-20C is not statistically significant in either ERA5 or NOAA 20CR. These comparisons suggest that the positive NAM–NAO association near zero lag is more consistent across the datasets, whereas the*

longer-lag relationships remain uncertain. Given the relatively short ERA5 record, the lack of upper-air observational constraints in CERA-20C and NOAA 20CR, and the non-stationary nature of the low-frequency variability, these results may be affected by sampling uncertainty and should be interpreted with caution."

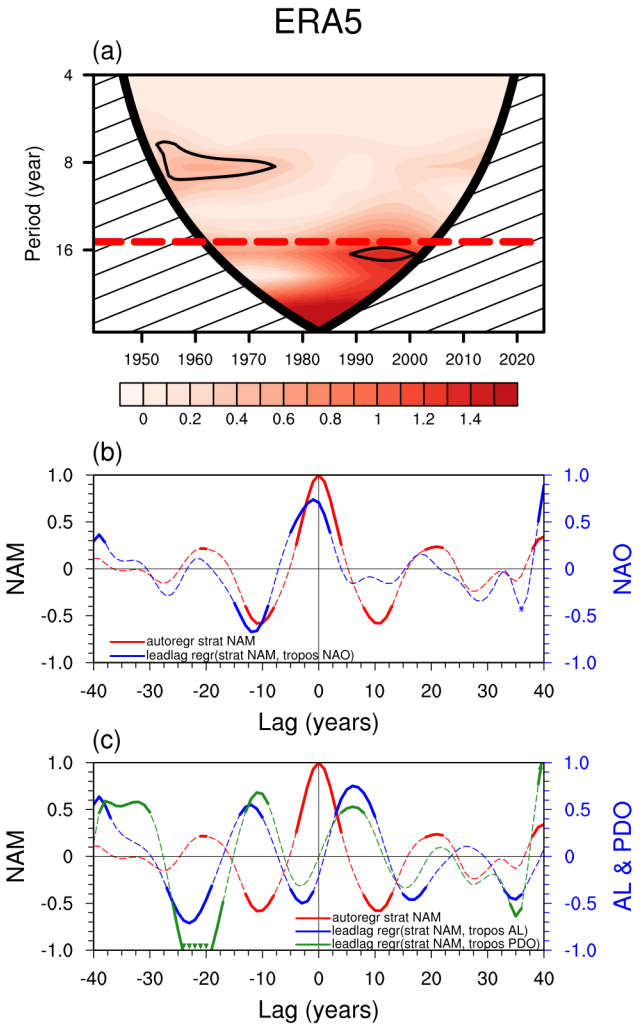


Figure R4. (a) Wavelet power spectrum of the stratospheric NAM index in ERA5. (b, c) Lagged regressions of the NAO, Aleutian Low and PDO onto the stratospheric NAM index.

4. I am also not convinced that the 15-years lagged SLP response in CERA-20C has a NAO signature (Figure 5a3). While it might project on the positive NAO, the actual response is dominated by a negative anomaly at around 50N where a NAO node is usually located. Since ability of the CERA-20C to simulate strat.-trop. coupling is not discussed, I am not sure if it is a

problem of CERA-20C or if there is an exaggerated NAO response in the models. This is one more reason to verify these SLP responses against those in ERA5.

Response: We thank the reviewer for this comment. We agree that the +15-year SLP regression pattern in CERA-20C does not exhibit a positive NAO dipole. Although the regression coefficient for the NAO index is positive at this lag (Figure 3), the SLP pattern in Figure 5 is dominated by a negative anomaly centered near 50°N, close to the typical NAO node. We have therefore revised the manuscript to describe the +15-year CERA-20C pattern more cautiously (**Lines 285-293**): "*At zero lag, positive NAM anomalies are associated with a positive NAO-like SLP dipole over the North Atlantic in CERA-20C, WACCM, and CAM6, consistent with the NAM–NAO regressions in Figure 3. However, at a lag of +15 years, the SLP pattern in CERA-20C does not exhibit a positive NAO dipole, despite the positive NAM-NAO regression coefficient in Figure 3. The pattern instead features a negative SLP anomaly centered near 50°N, close to the usual nodal line of the NAO SLP dipole.*"

As the reviewer suggests, we compare the lagged SLP regression patterns in ERA5 and CERA-20C (new Figure A6; shown here as Figure R5). At zero lag, both reanalyses show negative SLP anomalies over the high-latitude North Atlantic and positive anomalies farther south, resembling a positive NAO pattern. However, at +15 years, ERA5 does not reproduce the North Atlantic SLP pattern found in CERA-20C. The North Atlantic anomalies in ERA5 are weaker and have a different spatial distribution, with the strongest SLP anomalies occurring over the North Pacific. At −15 years, ERA5 exhibits a negative NAO-like pattern, whereas the corresponding pattern in CERA-20C is weak and less clearly defined.

These comparisons suggest that the NAO-like SLP pattern at zero lag is relatively consistent between the two reanalyses, whereas the lagged SLP patterns at ±15 years are less consistent. We have therefore revised the conclusion to emphasize that the NAM–NAO association is most evident near zero lag (**Line 365**): "*Lead–lag regression analysis suggests that stratospheric ~15-year variability is associated with the tropospheric NAO, particularly near zero lag.*"

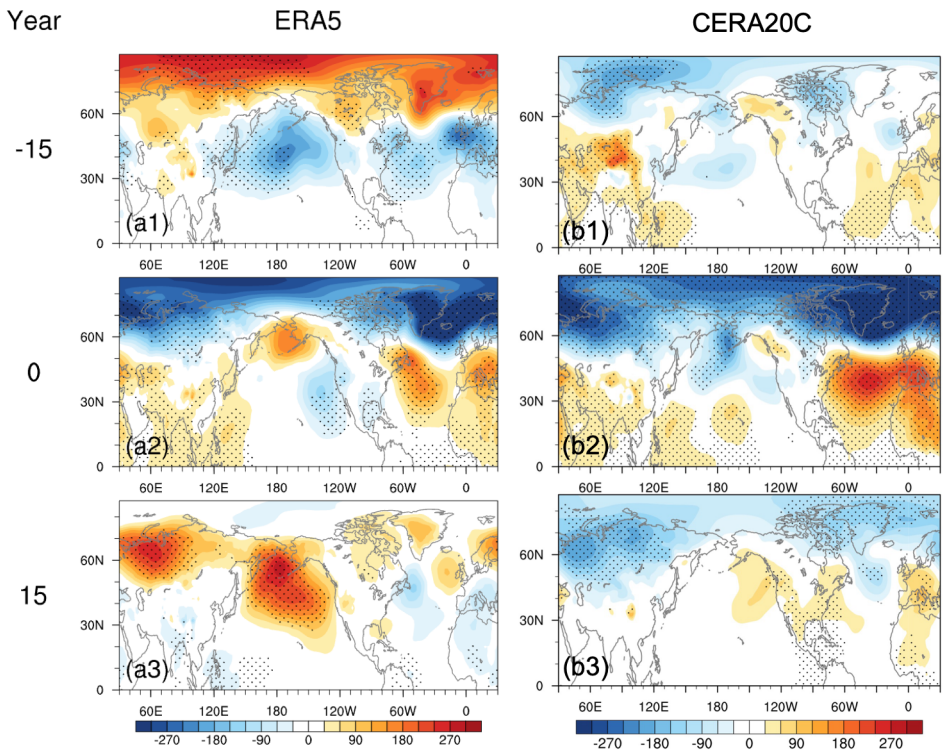


Figure R5. (a) Lagged regression patterns of SLP with the stratospheric NAM index for (a) ERA5 and (b) CERA-20C. (a) is shown as new Figure A6 in the revised manuscript. (b) is included in original Figure 5.

Few minor comments:

1. L210: "…while negative NAM anomalies are associated with negative NAO anomalies at lags near 22 years…" There is a negative correlation between NAM and NAO at lag +22. Therefore, I believe negative NAM anomalies should be associated with positive NAO anomalies.

Response: Thank you for pointing this out. The negative correlation represents the opposite relationship between positive NAM at lag 0 and negative NAO at lag +22. We have revised this sentence as: "*In CERA-20C, positive NAM anomalies at lag 0 are associated with positive NAO anomalies at lags near 0 and 15 years, and with negative NAO anomalies at lags near 22 years.*"

2. L288 "the negligible regression coefficient near zero lag". But they are statistically significant. Is it correct to call them "negligible"?

Response: We have replaced "negligible" with "*small in magnitude*".

3. Figure 3 (also 4 and 6): It would help if you mentioned in the captions that the y-axis for NAO is not only inverted but also its scale is different from that for NAM.

Response: Thank you for this suggestion. We have revised Figures 3 and 6 to remove the inversion of the NAO y-axis for clearer presentation. We have added the following sentence to the figure captions to clarify the different y-axis scales: "*The NAM and NAO indices are plotted against the left and right y-axes, respectively, with different scales*."