Replies to Reviewer #2:

Data from reanalysis and long climate-model simulations are used to investigate low-frequency variability in the high-latitude stratosphere on decadal to multidecadal timescales (~10–40 years) and its relationship with surface climate. Spectral analyses indicate significant stratospheric variability at a period of approximately 15 years in all data sets, together with additional spectral peaks at longer periods. The results also provide some evidence that stratospheric variability is followed approximately 15 years later by variability in the tropospheric North Atlantic Oscillation (NAO). No comparable stratosphere–troposphere relationship is identified over the North Pacific.

The manuscript addresses a topic of interest to the readership of Weather and Climate Dynamics and is well suited to the special issue on decadal climate variability. I also commend the authors for tackling a particularly challenging problem. The principal difficulties include the limited length of the observational record, the apparently red-noise-like and potentially nonstationary character of stratospheric variability, and the small signal-to-noise ratio. These limitations admittedly make it difficult to obtain robust and statistically significant results. Nevertheless, as detailed below, I have several concerns regarding the analysis and the interpretation of some of the results. These issues require attention before the manuscript can be considered for publication. I therefore recommend ***major revisions****.*

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

Major comments:

The main conclusions of the paper rely heavily on lagged regression analyses relating the stratospheric NAM to itself and to the North Atlantic NAO (Figs. 3 and 5) and the North Pacific Aleutian Low (AL; Figs. 4 and 5). However, the manuscript does not clearly explain the motivation for these analyses or how the resulting lag relationships should be interpreted. My

understanding is that the authors aim to determine whether the 15-year variability identified in the stratosphere induces variability on similar timescales in the troposphere, or whether the direction of influence is instead from the troposphere to the stratosphere.

The regression analysis of the reanalysis data indicates that NAM anomalies are associated with same-signed NAO anomalies both simultaneously and when the NAO lags the NAM by approximately 15 years. By contrast, no significant relationship is found when the NAO leads the NAM by 15 years. The authors interpret this asymmetry as evidence that the approximately 15-year low-frequency variability in the stratosphere excites corresponding variability in the troposphere.

However, my understanding is that the downward influence of the stratosphere on the troposphere operates on relatively short timescales, such that NAM and NAO anomalies should be approximately in phase. This would naturally explain the comparatively strong positive regression between the NAM and NAO at zero lag in Fig. 3a. By contrast, I find it difficult to understand the physical meaning of the positive association between the NAM and the NAO 15years later, or why the absence of a corresponding relationship when the NAO leads the NAM should be interpreted as evidence for a downward influence.

If the NAM exhibits quasi-oscillatory variability with a period of approximately 15 years and the NAO responds contemporaneously, positive regressions should occur at both positive and negative 15-year lags in an idealized stationary system. The lag direction alone therefore does not establish causality. Conversely, if the NAM variability is strongly intermittent and nonstationary, it remains unclear why a NAM anomaly should produce a similar NAO anomaly specifically 15 years later.

Moreover, an apparent asymmetry between positive and negative lags could result from the limited record length, nonstationarity, or sampling variability. In summary, the authors should explain much better the expected lag-regression structure under their proposed mechanism and clarify why the result should be interpreted as evidence that 15-year stratospheric variability drives the tropospheric NAO rather than as a consequence of the autocorrelation structure of the two indices.

Another major concern is the interpretation of the results. Many conclusions appear overly speculative given that the data are strongly smoothed and the relationships are often weak or only marginally significant. Moreover, the reanalysis and model simulations frequently show

different or even opposing signals. The discussion should focus more clearly on robust areas of agreement and disagreement, acknowledge these substantial inconsistencies and uncertainties, and avoid interpretations not directly supported by the results. In places, such as the discussion of Fig. A3, the text is also overly descriptive without explaining the physical significance of the patterns.

Response: We thank the reviewer for these constructive comments and suggestions. Our original motivation for the lead–lag regression analysis was to examine whether the stratospheric ~15-year variability is associated with variability in the tropospheric circulation and to explore the possible direction of influence between the stratosphere and troposphere. However, we agree with the reviewer that the lead–lag relationships alone do not provide sufficient evidence to establish causality.

As the reviewer pointed out, the positive NAM–NAO regression near +15 years does not necessarily indicate a delayed tropospheric response to stratospheric variability. For example, if positive NAM anomalies occur at lag 0 and again near +15 years due to the approximately 15-year variability, and the NAM and NAO are positively associated contemporaneously, positive NAO anomalies may also occur near +15 years, coinciding with the positive NAM anomalies at that time. We also agree that, in an idealized stationary system, positive regressions would be expected near both +15 and −15 years. The asymmetry in our results may therefore indeed be influenced by the limited record length, nonstationarity, and sampling variability.

Identifying the physical mechanisms and determining the direction of influence would require further dynamical analyses and targeted model experiments. We hope to explore these questions in future work. In this study, we mainly focus on the statistical relationships between stratospheric and tropospheric variability.

We agree that our original wording was too strong. We have therefore revised the Abstract, Main text, and Conclusions to remove wording implying that stratospheric variability leads or drives the NAO and instead emphasize only their low-frequency statistical association (**please see Lines 9, 220, 364**). We have also added a discussion discussing why the lead–lag relationships do not establish causality (**Lines 229-235**): "*These lead–lag relationships represent statistical associations and do not establish a causal direction between the NAM and the NAO. The positive NAM–NAO regression near +15 years may partly reflect the ~15-year autocorrelation of NAM combined with the positive contemporaneous NAM–NAO association.*

Positive NAM anomalies recurring near +15 years may coincide with positive NAO anomalies in the same winter, without implying a delayed tropospheric response to NAM anomalies at lag 0. The asymmetry between the regressions near +15 and −15 years may be influenced by the limited record length, the intermittency and nonstationarity, and sampling variability."

We would like to appreciate the reviewer's detailed explanation of why the lead–lag relationships alone cannot establish a causal influence of stratospheric variability on the NAO. These insights have been very helpful in improving our interpretation of the results. We sincerely thank the reviewer for the time and effort devoted to reviewing our manuscript.

To address the reviewer's second major concern regarding the interpretation of the results, we have revised the manuscript to describe the regression patterns more objectively, acknowledge the inconsistencies across datasets, and avoid overstating the evidence. The main revisions include:

(1) We have acknowledged the stronger positive NAM–NAO regression near −15 years in CAM6 and removed wording suggesting that the stratospheric NAM leads the NAO (**please see response to Comment 16**).

(2) We have revised the interpretation of the CERA-20C SLP pattern at +15 years, acknowledging that it does not exhibit a positive NAO dipole (**please see response to Comment 23**).

(3) We have revised the discussion of the North Atlantic SST patterns to describe the differences among CERA-20C, WACCM, and CAM6 more accurately. We have also clarified the implications of these differences, noting that the relationship between NAM variability and North Atlantic SST anomalies is not consistent across the datasets (**please see response to Comment 26, 27**).

(4) We have removed some overstated statements (**please see response to Comments 24, 32, 33**).

Further details of these revisions are provided in our point-by-point responses below.

Additional comments

1. 80: It would be interesting to include other 20th century reanalysis in this analysis, such as NOAA's 20CR.

Response: We thank the reviewer for this suggestion. We have repeated the wavelet and lead–lag correlation analyses using NOAA 20CR for 1837–2015 (new Figure A5; shown here as Figure R6).

The NOAA 20CR wavelet analysis shows enhanced power around 15-year timescales during the 1940s–1960s, partly overlapping with the ∼15-year signal in CERA-20C during the 1940s–1980s. These results show some similarities in the intermittent low-frequency stratospheric variability between the two reanalyses. NOAA 20CR also exhibits enhanced power near 30-year timescales.

We also examined the lead–lag regressions between stratospheric NAM and the NAO, Aleutian Low, and PDO indices. The NAM–NAO regression is significantly positive near zero lag, consistent with CERA-20C. However, unlike CERA-20C, NOAA 20CR does not show clear secondary peaks in NAM autoregression near ±15 years (red line), and the NAM–NAO regressions are also weak at these lags. Over the North Pacific, the NAM–Aleutian Low regression has the opposite sign to that in CERA-20C near zero lag, further illustrating the lack of consistency in the low-frequency Pacific relationship across the reanalyses and climate models.

We note that NOAA 20CR does not directly assimilate upper-air observations, so its stratospheric results should be interpreted with caution. We have included the results in the Supplementary Material (new Figure A5) and Discussion (**please see Line 406-412**).

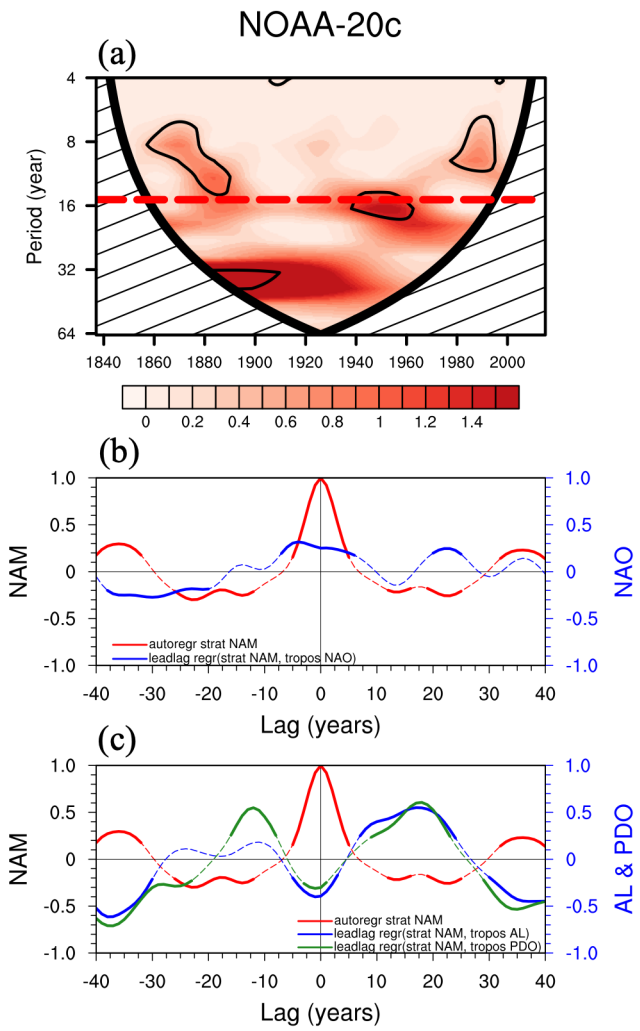


Figure R6. (a) Wavelet power spectrum of the stratospheric NAM index in NOAA 20CR. (b, c) Lagged regressions of the NAO, Aleutian Low and PDO onto the stratospheric NAM index in NOAA 20CR.

2. 110: The authors infer from 10 hPa Z that CAM4 has a stronger polar vortex than CAM6. This difference would be demonstrated more directly and clearly using the zonal-mean u structure. Also, it would also be informative to report the SSW frequencies in the different data sets.

Response: Thank you for this suggestion. We would like to clarify that our original statement at line 110 was intended to compare both CAM4 and CAM6 with CERA-20C, rather than to suggest that CAM4 has a stronger polar vortex than CAM6.

Following your suggestion, we have added four panels showing the climatological winter zonal-mean zonal wind structures for CRRA-20C and three models to the existing Figure A1 (**please see the fourth column in Figure A1**). These results show that WACCM has westerlies similar to those in CERA-20C, whereas CAM4 and CAM6 exhibit stronger westerlies.

We agree that comparing SSW frequencies would be informative. However, we use monthly mean output in this study. Monthly mean data are not suitable for reliably identifying individual SSW events and estimating their frequencies, as SSW identification typically requires daily zonal-mean zonal wind data.

3. 115-125: The authors derive NAM, NAO, AL, and PDO indices using EOF analysis and describe them as "normalized." Please define the normalization procedure explicitly. If each index is divided by its standard deviation, this removes differences in the absolute variability of these modes among datasets. This is particularly relevant if the magnitudes of regression coefficients are compared across datasets. The authors should also explain the EOF sign conventions, particularly what constitutes a positive AL or PDO phase. Showing the corresponding EOF patterns in physical units would clarify both the signs and amplitudes of these modes.

Response: Thank you for your suggestions. We have defined the normalization in the Methods section (**Line 127**): "*Each index is standardized by subtracting the mean from the corresponding principal component (PC) time series and dividing by the standard deviation.*"

We have also clarified the EOF sign conventions and added a new supplementary figure (new Figure A2; shown here as Figure R7) showing the spatial patterns associated with positive values of the four indices. We have added the following description to the Methods section (**Line 130-134**): "*The EOF signs are chosen such that positive NAM index indicates a stronger polar vortex; positive NAO index corresponds to the positive NAO phase, characterized by negative SLP anomalies over the subpolar North Atlantic and positive anomalies farther south; positive AL index indicates a deeper Aleutian Low, characterized by negative SLP anomalies over the North Pacific; positive PDO index corresponds to the positive PDO phase, characterized by*

negative SST anomalies over the central North Pacific and positive anomalies along the west coast of North America (Figure A2)."

We agree that standardizing each index removes differences in the absolute variability. Our lead–lag regression analysis based on standardized indices focuses on the statistical relationships between the stratospheric NAM and the surface circulation indices, but cannot directly quantify differences in the absolute variability of these modes across datasets. We have clarified this in the Method section (**Line 128-129**): "*Because all indices are standardized, the regression coefficients describe statistical relationships between standardized index anomalies but do not reflect differences in the absolute variability of the modes across datasets.*"

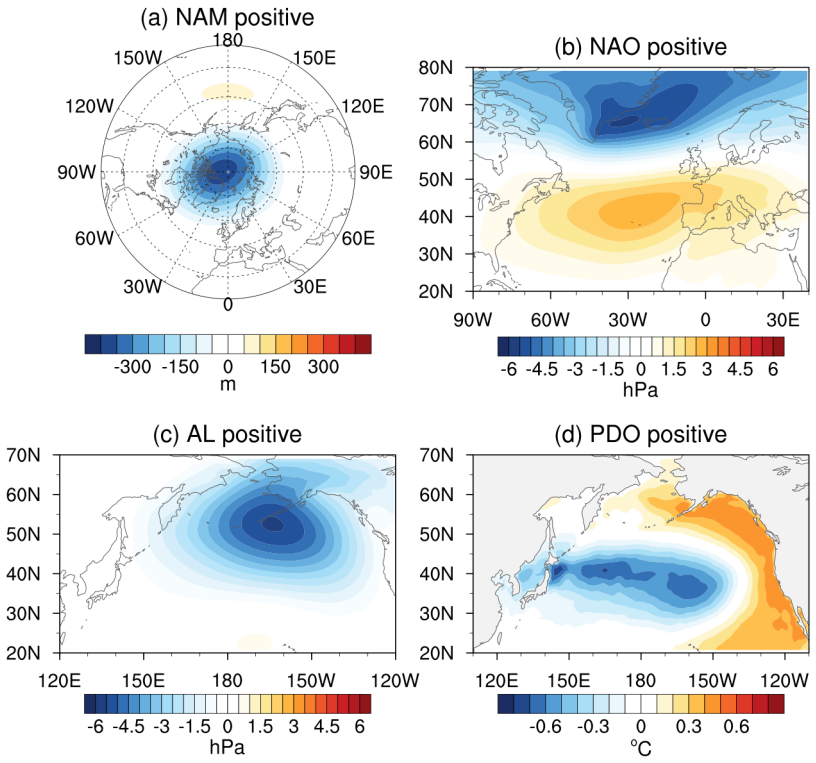


Figure R7. This is new Figure A2. Spatial regression patterns associated with positive (a) NAM, (b) NAO, (c) AL, and (d) PDO indices in CERA-20C. The patterns are obtained by regressing (a) 10-hPa geopotential height, (b, c) sea level pressure, and (d) sea surface temperature onto the corresponding standardized first principal component (PC1) time series.

4. 163: … the three CESM simulations …

Revised

5. 172: In the three model …

Revised

6. 173: … at 10-40 years …

Revised

7. Figure 1: The pronounced 15-year spectral peak in CERA is striking, but its physical origin remains unclear. Is this variability generated internally within the stratosphere, or does it involve coupling with the troposphere or ocean? The authors should discuss possible mechanisms and any evidence that distinguishes among them. Also, it would be interesting to see the corresponding spectra for the NAO and compare them with the NAM.

Response: We thank the reviewer for this suggestion. The physical origin of the ~15-year stratospheric variability remains unclear. Based on our current analyses, we cannot determine whether this variability arises from internal stratospheric processes or from interactions with the troposphere and ocean. Distinguishing among these possible mechanisms and establishing causality would require further targeted model experiments. We have therefore added a paragraph to the Discussion addressing possible sources of the ~15-year variability (**Lines 416–419**): "*There are still some unresolved questions. 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 modulation of the quasi-biennial oscillation (QBO). Internal stratospheric processes may also contribute.*"

Following your suggestion, we have calculated the NAO power spectra and wavelet spectra for ERA5, CERA-20C, NOAA-20C, WACCM, CAM4, and CAM6 (Figures R8 and R9). The NAO wavelet spectra show intermittent enhancements of power at periods near 15–20 years, particularly in ERA5, CERA-20C, and NOAA-20C (Figure R9). However, the NAO power spectra do not exhibit a consistently pronounced peak near 15 years across the datasets (Figure R8). These results suggest that the ~15-year spectral peak is more clearly expressed in the stratospheric NAM than in the tropospheric NAO. Nevertheless, these spectral comparisons

alone cannot establish a causal relationship between NAM and NAO or explain the physical origin of the ~15-year stratospheric variability.

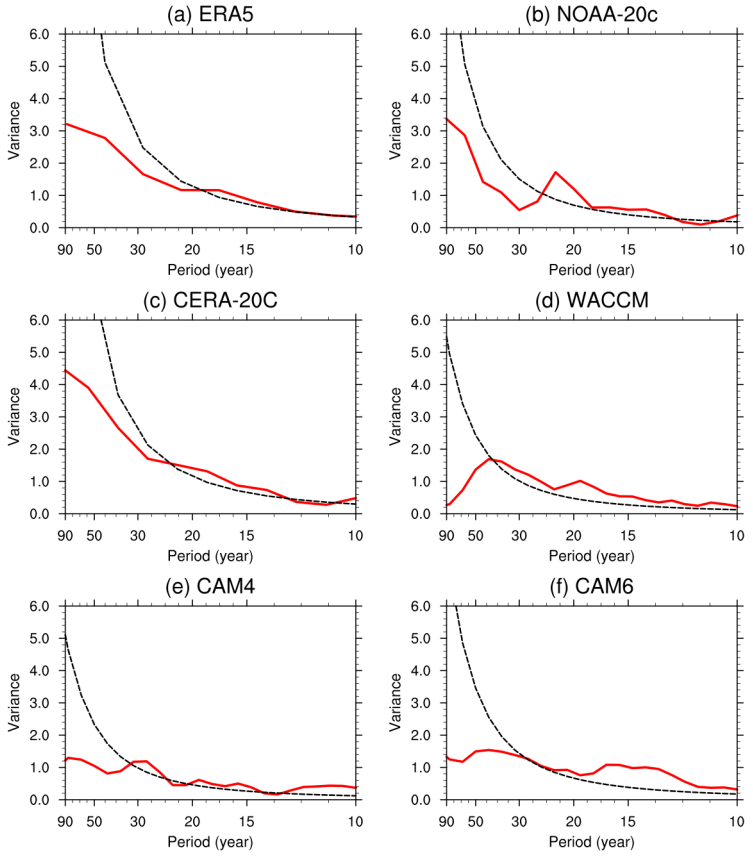


Figure R8. As in Figure 1, but showing the power spectra of the NAO index and additionally including ERA5 and NOAA-20C.

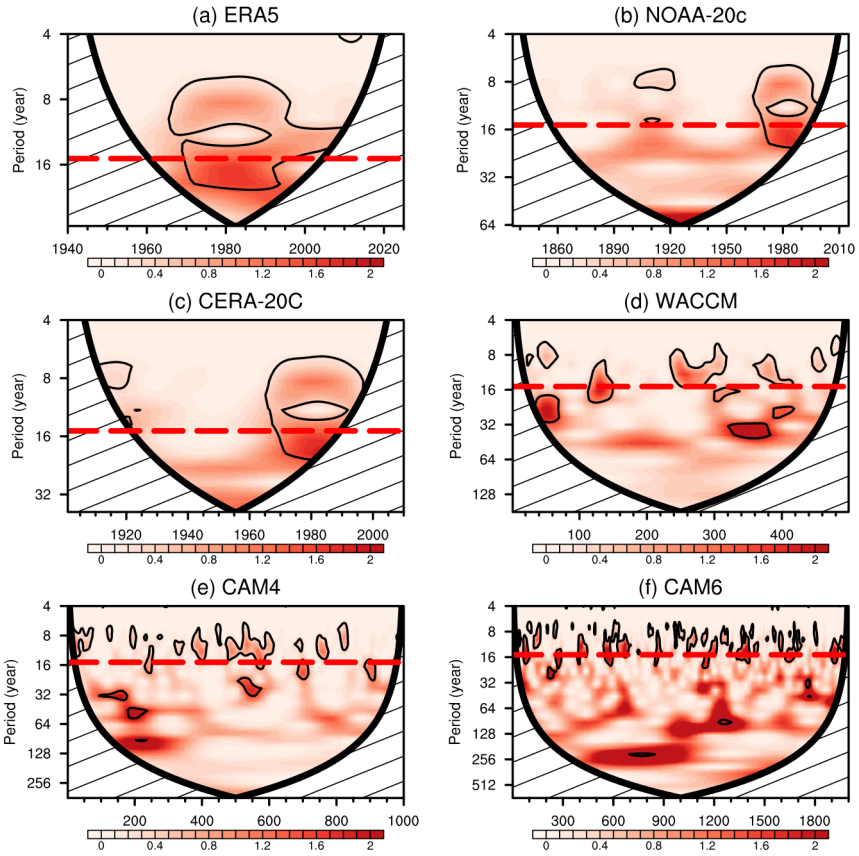


Figure R9. As in Figure 2, but showing the wavelet power spectra of the NAO index and additionally including ERA5 and NOAA-20C.

8. Figure 1: The spectral power near 15 years in WACCM appears relatively weak. This should be acknowledged and considered when interpreting the subsequent lag-regression results.

Response: Thank you for this suggestion. We agree that the relatively weak 15-year frequency peak should be acknowledged. We have revised the descriptions for WACCM (**Line 173**): "*Importantly, the same ~15-year peak is present in all the three CESM simulations, although the ~15-year spectral power is relatively weak in WACCM, with local peaks near 13–14 and 19 years (Figure 1b).*" We have also acknowledged this in the description in Figure 3 (**Line 213**): "*although the NAM autoregression near lags of ~15 years is relatively weak in WACCM (Figure 3b)*"

9. Figure 1: Why do you plot the CERA spectra at a different scale than the models?

Response: We have updated the new Figure 1 using the same y axis across CERA-20C and models.

10. Figure 1 caption: … from several 250-year segments …

Revised

11. 175: In the low-top …

Revised

12. 187: In a next step …

Revised

13. 200: It is not exactly clear what "focus on the ~15 year spectral peak" means for this analysis. Does it mean that the data are taken from segments that have large 15 year variability and little lower frequency variability, or is there any additional spectral filtering involved? I am asking this because the figure captions mention that the indices are smoothed with a Gaussian filter of sigma= 2 years, but the resulting Figs. 3-4 look very smooth.

Response: By "focus on the ~15-year spectral peak," we meant selecting time intervals with significant wavelet power near the 15-year period, rather than applying additional spectral filtering.

Specifically, we first smoothed the full time series of each index using a normalized, centered 21-point Gaussian kernel with $\sigma = 2$ years, following the approach used in previous studies of low-frequency stratospheric variability (Reichler et al., 2012; Dimdore-Miles et al., 2022). This Gaussian smoothing acts as a low-pass filter, reducing high-frequency variability while retaining lower-frequency variations. Based on the wavelet analysis, we then selected the time intervals with significant ~15-year variability for the lead–lag regression analysis, as specified in the figure captions. The relatively smooth regression curves result from the Gaussian smoothing applied to the original indices. No additional band-pass filtering or smoothing of the regression curves was applied.

14. 206: … but it does not …

Revised

15. 209: In Fig. 3a (blue line), the regression coefficient is positive near 0 and 15 years but negative near 22 years. A positive coefficient indicates same-signed NAM and NAO anomalies, whereas a negative coefficient indicates opposite-signed anomalies. Thus, the statement that "negative NAM anomalies are associated with negative NAO anomalies at lags near 22 years" appears incorrect; this would imply a positive regression. At 22 years, positive NAM anomalies should instead be associated with negative NAO anomalies, and vice versa. Similar language appears throughout the manuscript and should be revised. The authors should also assess whether correcting this interpretation affects any of their conclusions.

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 (**Line 223-224**): "*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*".

This clarification does not affect our conclusions. With positive NAM anomalies at lag 0 as the reference, the NAM autoregression indicates negative NAM anomalies near +22 years, and the NAM–NAO regression indicates negative NAO anomalies at the same lag. Our original wording therefore described the negative NAM and NAO anomalies at +22 years, rather than a relationship between negative NAM anomalies at lag 0 and negative NAO anomalies at +22 years. However, the reference time was not clearly stated. We have revised this sentence and similar wording throughout the manuscript to explicitly describe the lagged NAO, Aleutian Low, and PDO anomalies relative to positive NAM anomalies at lag 0. These revisions do not change the calculated regression coefficients or the signs of the anomalies at the corresponding lags.

16. 211: The statement that "CAM6 exhibits broadly consistent behavior" requires qualification. In CAM6, the positive regression near −15 years appears stronger and more significant than that near +15 years. Taken at face value, would this imply that the NAO leads the NAM by approximately 15 years? Although this question is partly rhetorical, it illustrates a broader

concern: results that support the proposed interpretation are discussed, whereas equally prominent features that do not fit that interpretation are often left unexplained. The authors should discuss the results more objectively and clarify which features are physically meaningful and which may arise from sampling variability.

Response: Thank you for pointing this out. We agree that our original description was too strong and did not acknowledge the stronger positive regression near −15 years in CAM6.

We have revised this paragraph to describe the results more objectively. First, we revised the opening sentence to avoid suggesting that the stratospheric NAM tends to lead the NAO (**Line 219**): "*CERA-20C, WACCM, and CAM6 suggest a low-frequency statistical association between stratospheric variability on ∼15-year timescales and NAO variability on similar timescales*". Second, we have explicitly acknowledged the stronger negative-lag signal in CAM6 (**Line 227**): "*However, CAM6 shows a more pronounced positive regression peak near −15 years than near +15 years, whereas the regressions near −15 years are weak and not statistically significant in CERA-20C and WACCM.*". Finally, we have recognized that the current analyses cannot determine causality (**Line 229**): "*These lead–lag relationships represent statistical associations and do not establish a causal direction between NAM and NAO.*"

17. 223: CAM4 shows a significant negative regression peak at a lag of approximately +5 years. How should this feature be interpreted in terms of strat-trop coupling?

Response: This indicates that positive NAM anomalies at lag 0 are associated with negative NAO anomalies approximately five years later. However, this feature is not consistently reproduced in the other datasets, and the physical mechanism remains unclear. We have added a sentence to acknowledge this (**Line 240**): "*CAM4 also shows a significant negative NAM–NAO regression near a lag of +5 years. This feature is not evident in the other datasets, and the physical mechanism remains unclear.*"

18. 227: See me comment on line 209.

Response: We have checked the sign and revised this sentence as: "*CAM6 continues to show a significant positive NAO anomaly at lags near 0 and 15 years following positive NAM anomalies at lag 0.*"

19. 231: The authors assess possible stratospheric influences over the North Pacific using both the Aleutian Low (AL) and PDO indices, but provide little explanation for their relevance or selection. Because the AL and PDO are closely related, it is unclear what independent information is gained by analyzing both. The similar results for CERA in Fig. 4a appear to support this concern. Please clarify the physical distinction between the two indices and justify including both in the analysis.

Response: Thank you for this comment. We included the PDO index to examine the statistical relationship between stratospheric NAM variability and North Pacific SST variability, in addition to the atmospheric SLP variability represented by the AL index. We agree that the AL and PDO indices are closely related and do not interpret their similar regression results as independent evidence of stratospheric influence.

20. 233: See me comments to lines 209.

Response: Thank you for this comment. We have checked the regression signs and revised the description to clarify the Aleutian Low and PDO anomalies associated with positive NAM anomalies at lag 0. The corrected Figure 4a shows positive NAM–Aleutian Low and NAM–PDO regression coefficients at zero lag, indicating that positive NAM anomalies are associated with a deepened Aleutian Low and positive PDO anomalies in CERA-20C. We have revised the corresponding text (**Line 257-262**): "*Near zero lag, positive NAM anomalies are associated with a deepened Aleutian Low in CERA-20C (Figure 4a). In contrast, the three climate models generally show the opposite relationship, with positive NAM anomalies associated with a weakened Aleutian Low (Figures 4b–d).*" This correction does not change our conclusion that the NAM–Aleutian Low relationship differs between CERA-20C and the climate models.

21. 236: How do the authors interpret the absence of a relationship in CERA at ±15 years? Is this consistent with the generally weaker stratosphere–troposphere coupling over the North Pacific?

Response: Thank you for this comment. The absence of significant NAM–AL and NAM–PDO regressions near ±15 years in CERA-20C might be related to the generally weaker stratosphere–troposphere coupling over the North Pacific. However, our current analyses cannot determine whether this is the reason for the lack of significant relationships at these lags.

22. 237: The WACCM regressions are largely not statistically significant and should therefore not be interpreted.

Response: We have removed the original interpretation of the WACCM regression pattern and revised the sentence as "*In WACCM, the NAM–Aleutian Low regression is significant only at a few isolated lags (around −18, −12, +2, and +25 years)*"

23. 255: I do not agree with the interpretation of Fig. 5. At +15 years, CERA does not show a clear positive NAO signal.

Response: 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 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 289–293**): "*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.*"

24. 258-259: This is a strong statement, but the CERA results in Fig. 5 do not corroborate those in Fig. 3.

Response: We agree that Figures 3 and 5 are not fully consistent at a lag of +15 years. Although Figure 3 shows a positive NAM–NAO regression coefficient, the corresponding SLP pattern in CERA-20C in Figure 5 does not exhibit a positive NAO dipole. We have therefore removed the original statement and explicitly acknowledged this difference as: "*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.*"

25. 261: Please begin a new paragraph before "In CERA," as the discussion shifts to SSTs.

Response: We have revised the manuscript to begin a new paragraph before "In CERA-20C," as suggested.

26. 262: In Fig. A3a, I do not see a clear tripolar SST pattern in CERA, particularly at +15 years.

Response: Thank you for pointing this out. At zero lag, CERA-20C exhibits a tripole-like SST pattern, with negative anomalies over the subpolar North Atlantic between approximately 50°–65°N, positive anomalies off the East Coast of North America, and a second negative anomaly off the coast of northwestern Africa. We agree that the SST pattern in CERA-20C does not exhibit a clear tripole structure at lag of +15 years. We have revised the paragraph to remove the description of a tripole pattern at +15 years and added the following clarification (**Line 302-303**): "*At lag of +15 years, the tripole-like structure is less evident, with positive SST anomalies primarily concentrated in the western and central North Atlantic.*"

27. 267: This interpretation appears overstated, as the CERA and WACCM SST patterns show little similarity. Similarly, stating only that CAM6 "exhibits a different SST pattern" is too descriptive.

Response: Thank you for this comment. We have removed the statement that WACCM reproduces a similar tripole pattern to CERA-20C. Instead, we now clarify that although WACCM exhibits a tripole-like SST pattern at zero lag, the two negative anomaly centers differ in location from those in CERA-20C and are not statistically significant (**please see Line 303-305**).

We also revised the description to clarify that CAM6 does not exhibit a clear tripole-like SST pattern over the North Atlantic at zero lag, and that the North Atlantic SST patterns in WACCM and CAM6 differ from those in CERA-20C at lags of ±15 years (**please see Line 305-307**). Finally, we added a sentence to conclude the SST results: "*These differences suggest that the relationship between NAM ~15-year variability and North Atlantic SST anomalies is not consistent across the reanalysis and climate models.*"

28. 269 and 271: I am confused: Fig. 5 does not show SST regressions. Are the authors still referring to Fig. 5 here, or should this be Fig. 4?

Response: Figure 5 shows the SLP regression patterns, whereas Figure A4 shows the SST regression patterns. We have added the corresponding figure references in the revised manuscript to make this distinction clearer (**Line 310**): "*Over the North Pacific, the SLP (Figure 5) and SST (Figure A4) regression patterns....*"

29. 274: Given the interpretation at line 233, Fig. 4 indicates a negative association of the NAM with the AL and PDO at zero lag in both CERA and WACCM. However, Fig. 5 shows opposite-signed SLP anomalies over the Aleutian Low region. How can these results be reconciled?

Response: We apologize for the error in the original Figure 4a, where the signs of the Aleutian Low and PDO regression coefficients were inadvertently reversed. We have corrected this error in the revised Figure 4a. The corrected results show that positive NAM anomalies are associated with a deepened Aleutian Low at zero lag in CERA-20C, consistent with the negative SLP anomalies over the North Pacific in Figure 5a2. We have also revised the corresponding description (**Line 257**): "*Near zero lag, positive NAM anomalies are associated with a deepened Aleutian Low in CERA-20C (Figure 4a)*"

This correction does not change our main conclusion that the relationship between the stratospheric ~15-year variability and North Pacific surface circulation is not consistent between the reanalysis and climate models.

30. 285: From Fig. 6, at 30–40-year lags, the WACCM NAM autoregression is not significant and the CAM6 autoregression changes sign. Is further interpretation therefore justified?

Response: Thank you for this comment. We would like to clarify that the ~30–40-year range refers to characteristic periods closer to 40 years in WACCM and around 30 years in CAM4 and CAM6, as indicated by the wavelet analysis (Figure 2). In WACCM, the NAM autocorrelation is positive and statistically significant near a lag of +40 years, as indicated by the solid red line in Figure 6a. In CAM6, the NAM autocorrelation is positive near +30 years (Figure 6c). Although

the CAM6 autocorrelation turns negative at longer lags toward +40 years, this sign change does not contradict a characteristic period of approximately 30 years.

Our original wording grouped WACCM and CAM6 together under the ~30–40-year range without clearly specifying that our interpretation focuses on lags near +40 years in WACCM and +30 years in CAM6. To avoid this ambiguity, we have made several revisions to Section 3.3:

(1) We changed the subsection title from "Stratospheric ~30-40-year spectral peaks" to "Stratospheric multidecadal spectral peaks". **(2)** In the first paragraph, we added "particularly near 40 years" to the description of the WACCM spectral peak to clarify our focus on the ~40-year timescale in WACCM. **(3)** In the second paragraph, we revised the discussion of Figure 6 to explicitly describe the NAM–NAO regression results at +40 years in WACCM and +30 years in CAM6 (**Line 331-334**): "*In particular, both models show positive NAM–NAO regression coefficients near zero lag. The NAM autocorrelation and NAM–NAO regression coefficient are positive near a lag of +40 years in WACCM (Figure 6a) and near a lag of +30 years in CAM6 (Figure 6c).*" **(4)** In the third paragraph, we use "multidecadal spectral peaks" when discussing the results in WACCM, CAM6 and CAM4 together, rather than referring to a single ~30–40-year peak.

31. 286: At 30–40-year lags, the NAM–NAO relationships in WACCM and CAM6 have opposite signs. This should be acknowledged and raises concerns about the robustness of the inferred relationship.

Response: Thank you for this comment. Please see our response to the previous comment for a detailed discussion of the different characteristic periods in WACCM and CAM6. Our interpretation focuses on lags near +40 years in WACCM and +30 years in CAM6, where the NAM autocorrelation and NAM–NAO regression coefficients are positive (Figures 6a and 6c). We have revised Section 3.3 to clarify these results (**please see our response to comment 30**). Although the detailed lead–lag relationships differ between the models, the results suggest a possible association between stratospheric multidecadal variability and the NAO.

32. 316: The phrase "is closely associated with" overstates the evidence. CERA shows no clear NAO-like response at a 15-year lag in Fig. 5, and the corresponding regressions in CERA and the models in Fig. 3 are only barley significant.

Response: We have removed "closely" from "is closely associated with" in the Abstract, main text, and Discussion to avoid overstating the relationship.

33. 327: Given the concerns above, describing CAM6 as "far more realistic than CAM4" is too strong and implies a level of realism not demonstrated here. "More consistent with CERA than CAM4" would be more appropriate.

Response: Thank you for this suggestion. We have revised the sentence as "*more consistent with CERA-20C than CAM4*"

34. Fig. 3 caption: The manuscript states that solid line segments indicate regression coefficients significant at the 95% level based on a Student's t-test, but the test procedure is not described. Please specify the degrees of freedom used and whether the effective sample size accounts for temporal smoothing and autocorrelation, as well as the decreasing number of overlapping observations at increasing lags.

Response: Thank you for this comment. The statistical significance of the regression coefficients in Figure 3 was assessed using a two-sided Student's t-test with (n-2) degrees of freedom, where (n) is the number of overlapping observations at each lag. The indices were smoothed using a 2-year Gaussian filter, and we did not adjust the degrees of freedom for temporal autocorrelation. Smoothing method increases autocorrelation and can substantially reduce the effective sample size, which is particularly challenging for the relatively short CERA-20C record (~110 years). We have acknowledged this limitation in the revised Figure 3 caption.

35. Fig. 4 caption: values represents > values represent

Revised