Review of Wang et al. 2026, WCD

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**.

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 15 years 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.

Additional comments

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

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.

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.

163: … <u>the</u> three CESM simulations …

172: In the three model …

173: … at 10-40 years …

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.

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.

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

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

175: In the low-top …

187: In a next step …

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.

206: … but it does not …

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.

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.

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?

227: See me comment on line 209.

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.

233: See me comments to lines 209.

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?

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

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

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

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

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

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.

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?

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?

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?

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.

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.

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.

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.

Fig. 4 caption: values represents > values represent