<u>Response to RC2:</u>

This study assesses changes in large-scale modes of climatic variability, such as El Nino/Southern Oscillation (ENSO) under scenarios of feedback controlled stratospheric aerosol injection (SAI), as modelled in two versions of the Community Earth System Model (CESM). The study addresses an important research question, about which there has been little research and which I believe is of interest to the community. However, I would not recommend publication of the study in its current form, as there are several important revisions required to demonstrate that the results of this paper are robust, and to bring clarity to the writing and figures.

<u>Reply</u>: We sincerely appreciate your effort and time in reviewing our manuscript as well as your constructive comments/suggestions. We have attempted to address all your suggestions/comments. The following is the point–point response to all the comments (comments are rewritten in black color and their corresponding replies in red).

- Firstly, changes in values of climate metrics are consistently interpreted as representing a forced response to SAI and/or warming without reference to significance testing, and without placing the magnitude of changes in the context of internal variability. For example, Figure 1(l) is described as showing that "SAI in CESM2 effectively restores the projected changes [in the PDO]" (line 223-224), but it is not clear from the figure that this is the case. There is a slight increase in the median, towards it's historical value, under SAI, but the distribution as a whole as represented by the box and whiskers appears to see a decrease. No statistics are given from which to judge the significance of the change. Similarly, Figure 6 is described in the text (around lines 294-300), and implicitly by the red arrows marked on the figure, as showing changes in metrics of El Nino and La Nina, based on changes in medians between scenarios, some of which are very small and many of which sit well within the shown interquartile ranges. No quantification of the statistical distinguishability of the distributions behind the box plots is given.

<u>Reply</u>: Implemented. We have conducted a t-test analysis on the SST variances in the new Fig. 2 (old Fig. 1) and the El Nino/La Nina characteristics in the new Fig. 7 (old Fig. 6). Upon your suggestion in the last major comment, we further re-plotted Fig. 2 to only consider those elements of CECM1 contain the longer data since the longer series, the more reliable periodicity changes. The result descriptions for Figs. 2 and 7 have been re-written as follows:

"For RCP8.5 and SSP5-85 using CESM1 and CESM2, respectively, the strong greenhouse gas forcing and global warming to the end of the 21st century increases the variance of the first EOF SST anomaly in the North Atlantic and North Pacific (representing AMO and PDO), but reduces the variance of the leading EOF in global SST anomaly (related to ENSO). Based on the statistical t-test results, the changes in the means imposed by global warming relative to historical are all significant except one case (Fig. 2f). Differences between SAI and historic in CESM2 values of the leading EOF variance of AMO and ENSO are not significant, showing that the significant changes under greenhouse gas forcing are effectively reversed by SAI. In contrast the significant changes in PDO variance imposed by global warming using CESM1 relative to historical remain significant under SAI. Using CESM2, there is no significant changes in the PDO variance from historical to global warming, or to SAI."

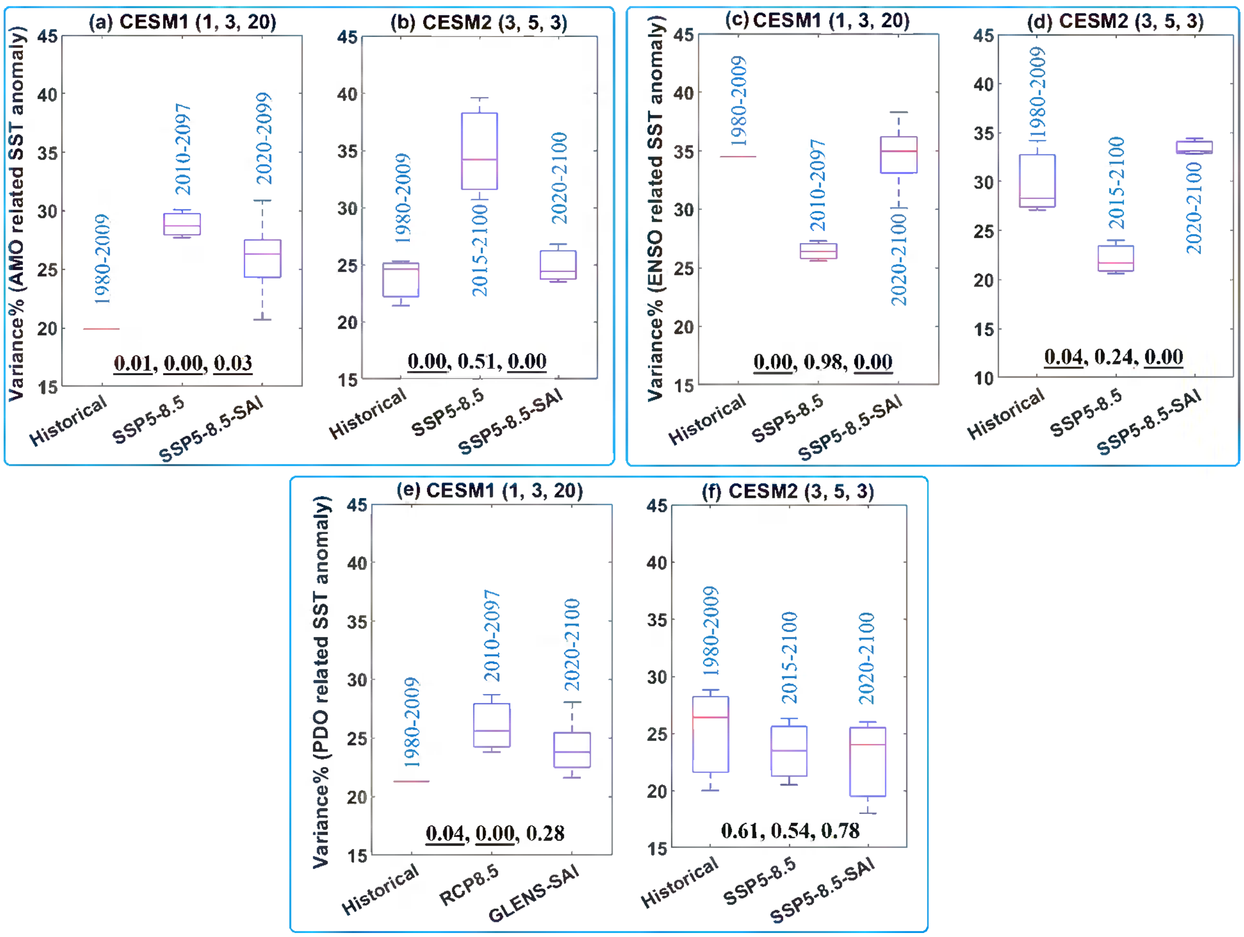

Figure 2. Box and whiskers plot of the variance in the leading EOFs, representing AMO, PDO, and ENSO, relative to the total variance of the SST fields: AMO across the North Atlantic (top-left panel); ENSO (top-right panel) global SST; and PDO across the North Pacific (bottom row). The values in blue on each column box show the period of the data for historical, greenhouse gas (i.e., RCP-8.5 and SSP5-8.5), and climate intervention (GLENS-SAI and SSP5-8.5-SAI) scenarios. The titles of each subplot refer to the CESM version and the number of ensembles used in the historical, greenhouse gas, and SAI scenarios, respectively. The median for each experiment is denoted by the red line, the upper (75th) and lower (25th) quartiles by the top and bottom of the box and ensemble limits by the whisker extents. The three values shown at bottom of each sub-plot refer to the p-values obtained

from the statistical t-test between historical and global warming, historical and SAI, and global warming and SAI, respectively. Values underlined are significant (i.e., $p<0.05$).

"Figure 7 displays the projected changes in the El Niño and La Niña episodes in the ENSO index under global warming and SAI. The global warming scenario simulated by CEMS2 tends to reduce the time between, as well as the intensity and duration of the La Niña episodes compared to the historical conditions, but El Niño shows no significant changes. Frequency increases in both El Niño and La Niña episodes were suggested in earlier climate simulations e.g., Fredriksen et al. (2020), Cai et al. (2014) and Yun et al. (2021) for El Niño, and Cai et al. (2015) for La Niña. In contrast, using CESM1, the characteristic changes of El Niño are stronger than that of La Niño and the El Niño intensity significantly increases while its duration decreases relative to historical period. The La Niño intensity significantly increases but other characteristics show no significant changes under RCP8.5.

Although the SAI is mostly accompanied by a slight decrease in the median of El Niño/La Niña characteristics towards their historical value, its effect on global warming imposed-changes is only statistically significant for the intensity and duration of La Niña events. For the CESM2 SAI experiment, there are no significant differences in El Niño characteristics as with the greenhouse gas forcing experiment. In contrast La Niña peak intervals, height (i.e., intensity), and width (i.e., duration) characteristics are significantly different from greenhouse gas forcing and reverse the direction of changes imposed by greenhouse gases. For CESM1, there are no significant differences between the results from RCP8.5 and GLENS-SAI scenarios."

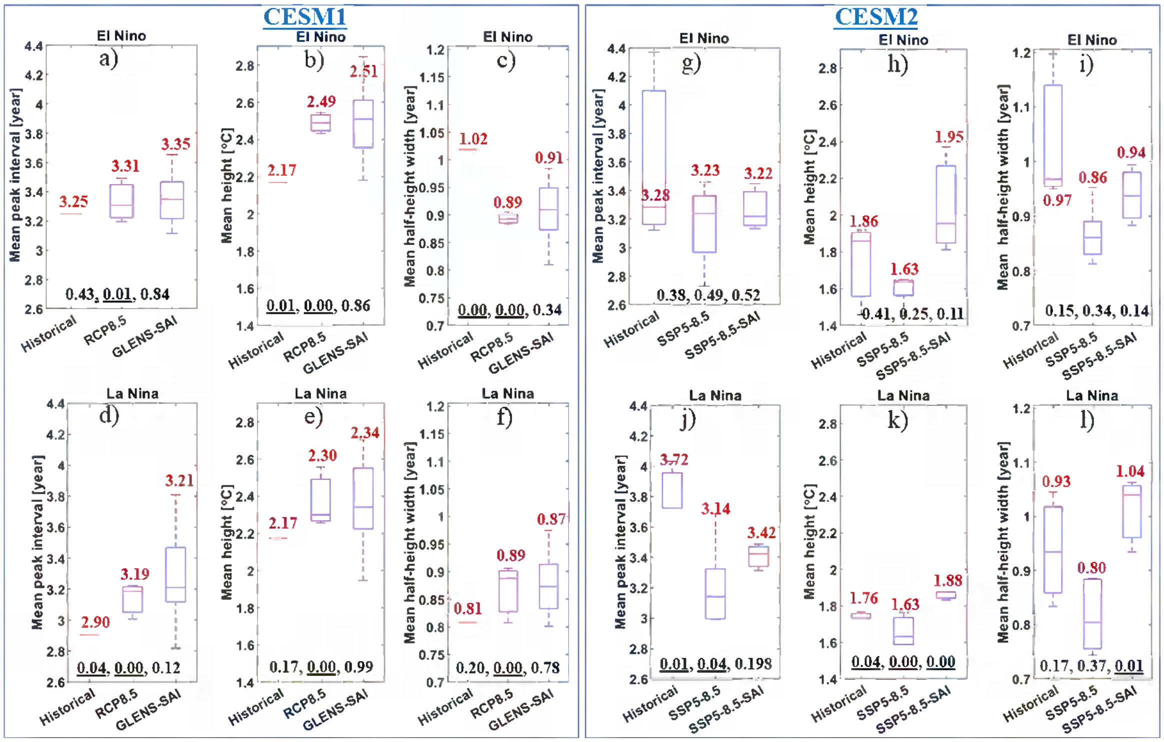

Figure 7. The projected changes in the mean peak interval, height, and half-height width of El Niño and La Niña events for global warming (RCP8.5 and SSP5-8.5) and SAI (GLENS-SAI and SSP5-8.5-SAI) scenarios simulated by CESM1 (panels a-f) and CESM2 (panels g-l). The median for each experiment is denoted by the red line, the upper (75th) and lower (25th) quartiles by the top and bottom of the box, and ensemble limits by the whisker extents. The values labeled in red on each box show their median. The three values shown at bottom of each sub-plot refer to the p-values obtained from the statistical t-test between historical and global warming, historical and SAI, and global warming and SAI, respectively. Values underlined are significant (i.e., $p<0.05$).

- (a) This study mentions that in the only previous assessment of ENSO under SAI, by Gabriel and Robock (2015), SAI simulations may not have been long enough to detect changes. I assume that the large 20-member ensemble of GLENS may overcome this limitation, especially for short-period indices, since this represents ~1600 model-years. (b) The authors should consider adding a discussion of whether this is the case, and for each index, the size of change which their analysis would allow them to detect. (c) Perhaps the pre-industrial run for each model could be added to the analysis and used to characterize the tendency of each of these indices to vary on the timescales considered, to contextualise the scale of changes shown in figures 1 and 6. (d) Some discussion of multiple testing is also necessary, since multiple indices, with multiple features of each are assessed. (e) In the absence of any theoretical argument for why we might expect particular changes in these indices under SAI, there is a danger of cherry picking the largest changes amongst many noisy time series.

Reply: (a) ~1600 model-years: It should not work to concatenate all the ensemble members into a single long times series since the phases of the long period signals will be different in each ensemble member - assuming that each member was branched sufficiently far before the analysis period. The act of concatenating the individual time series will mean that the component signals within them will be placed discontinuously in phase, and the amount of discontinuity will vary according to the period. Nonetheless, we analyzed the concatenated series from the available members for each scenario using CESM2, excluding CESM1 outputs as there is solely a single ensemble member for CESM1 historical data over a short 1980-2009 period. Figure S5 summarizes the CWT global power spectrums for AMO, NAO, ENSO, and PDO. The results, on the whole, are compatible with those shown in Fig. 8 (below in the response to minor comment 2), despite small discrepancies such as the very stronger interdecadal mode in AMO obtained from the concatenated ensembles.

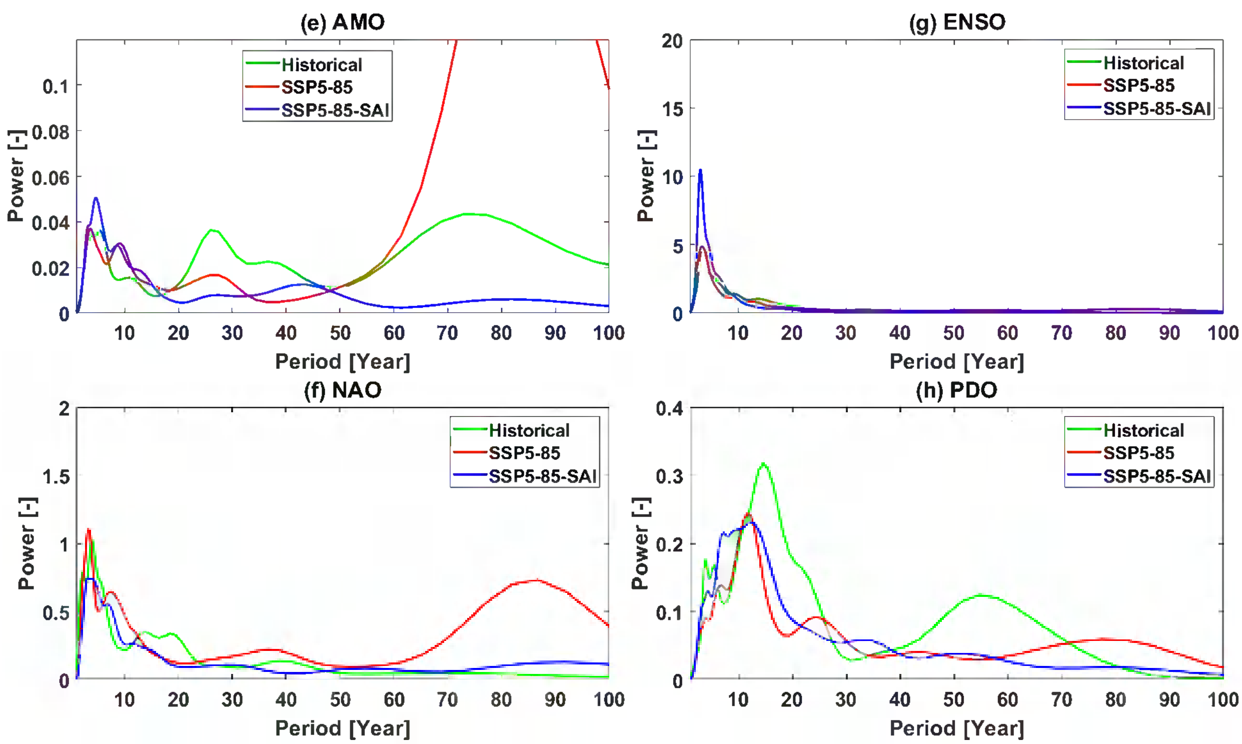

Figure S5. As Fig. 8 but obtained from the concatenated members of the CESM2 historical simulations.

(b) the size of change detectable: The wavelet method cone of influence (COI) automatically shows where the periods analyzed are being influenced by the end of the time series. Thus, the longest periods can only be reliably assessed for the middle of the time series. The changes detectable are determined by significance tests: Monte Carlo methods in the WTC case and t-tests for other changes.

(c) pre-industrial run for each model: The pre-industrial analyses we use now span the full range of the times series available. Concatenating historic and SAI will produce continuous results at longer periods than we can now analyze, but they will still be a mix of two different climate states (pre-industrial and SAI).

(d) multiple testing: Please refer to the response to the first major comment above. We have conducted a t-test analysis on the SST variances in the new Fig. 2 and the El Nino/La Nina characteristics in the new Fig. 7.

(e) cherry picking: We have conducted a t-test analysis, please refer to the response to the first major comment above.

- (a) More clarity would be useful over how the authors intend to treat agreement and disagreement between the two models and their SAI simulations. It is suggested that (line 120) the two members of the CESM family are different enough to explore 'a range of plausibly real climate impacts'. (b) There is, however, no discussion of how these two members of the CESM family compare to the inter-model spread in representation of these climate indices in, e.g., the CMIP6 ensemble, and to observations. (c) More models could be added using GeoMIP simulations, should the authors wish to do so (albeit for a different SAI scenario).

Reply: We have added more explanations to the text body of the paper as follows:

a) two members of the CESM family are different enough: The SAI scenarios using both CESM1 and CESM2 inject SO_2 at four predefined points (30°N, 30°S, 15°N, and 15°S) at ~5 km above the tropopause using a feedback controller to maintain not just the global mean temperature, but the interhemispheric and equator-to-pole temperature gradients. Fasullo and Richter (2022) explain the inter-model differences in the aerosol mass latitudinal distributions between the SAI experiments using CESM1 and CESM2. CESM2 SAI utilizes the CMIP6 SSP5-8.5 experiment as a baseline which has been used by various modeling teams (Tilmes et al., 2020) while CESM1 SAI uses the well-known RCP8.5 scenario. In GLENS-SAI, most of the aerosols were injected at 30°N and 30°S with much smaller injection mass at 15°N and a tiny amount at 15°S while for SSP5-8.5-SAI, the highest concentrations were released at 15°S, modest mass at 15°N and 30°S, and a small amount at 30°N. These differences in the SO_2 distributions across the two SAI scenarios for CESM1 and CESM2 produce a range of variability in shortwave radiation and cloud responses to CO_2 concentration increases (Fasullo and Richter, 2022). Additionally, Fasullo and Richter (2022) identified that changes in the spatial salinity and density patterns in the Atlantic Ocean, and in turn, the Atlantic Meridional Overturning Circulation (AMOC), are very different under GLENS-SAI compared to SSP5-

8.5-SAI experiment. These differences between SAI simulations represent part of the system variability.

b) compare to the inter-model spread: The equilibrium climate sensitivity (ECS) of CESM2-WACCM is 4.75 °C and lies in an ECS range of 1.83 to 5.67 °C from 41 different CMIP6 GCMs (IPCC AR6, 2021). The absolute mean surface temperature difference between CESM2-WACCM and historical records (0.89 °C) and is also within the range of 0.38-1.23 °C from 37 different CMIP6 models (Scafetta, 2021). CESM2 is one of the best nine models for simulating precipitation worldwide when measured by the Hellinger distance between bivariate empirical densities of 34 CMIP6 models and the historical data from Global Precipitation Climatology Centre (GPCC; Abdelmoaty et al., 2021). Additionally, the global-mean values of SST, summer land temperatures, precipitation, and ECS simulated by CESM1 and CESM2 are roughly similar to each other as well as compatible with the historical values over the 1985-2014 period (Danabasoglu et al., 2020; Table S1).

Relative to the preindustrial 1851-1850 period, CESM2-WACCM projects global mean surface air temperature rises of ~6.25 °C by the 2071-2100 period under SSP5-8.5 which compares with the range of ~3.3-6.6 °C from 35 ensembles of 12 CMIP6 models (Cook et al., 2020).

Table S1. The global-mean values of SST, JJA land temperature minus historical, precipitation, and ECS simulated by CESM1 and CESM2 relative to the historical values over the 1985-2014 period (modified after Danabasoglu et al., 2020).

Variable	SST (°C) over 1985-2014	JJA land T (°C) minus historical (1985-2014)	Precipitation(mm/d) over 1985-2014	ECS (°C)
Historical	18.29	-	2.69	1.5 – 4.5
CESM1	18.02	-0.99	3.02	4.0
CESM2-WACCAM	18.63	0.25	2.92	5.1
CESM2-CAM6	18.69	0.37	2.94	5.3

(c) More models could be added using GeoMIP: The referee refers to different models running the GeoMIP G6 which exist now. We chose not to use them in this analysis because we wanted as large a signal from SAI as possible. The changes in the teleconnection patterns are subtle, and the figures are complex and not easy to interpret. The G6 scenarios do not aim to return the climate to the historical, but to lower radiative forcing from extreme to moderate levels, thus making examination of changes relative to historic impossible, and also making a simple comparison with what we refer to here simply as greenhouse gas or global warming more complicated.

- The findings of suppressed long-period variability in the AMO under SAI relative to both historical and warming, and the un-restored long-period variability in the PDO under SAI, in CESM2 are perhaps the starkest changes seen, and worth more discussion. However, they are found only a 3-member ensemble for one model, and as such, the authors should consider more strongly caveating their statements, particularly in the abstract (line 33).

<u>Reply</u>: We have added more caveats as follows to the text:

"However, these findings obtained from indices of 85–90-year length for global warming and ~80-year for SAI, while long enough to explore variability within the statistical assumptions of the methods may not be robust against non-stationary time series where the assumptions inherent in the wavelet and e.g., t-tests would not hold. We are limited to the available simulations, and a 3-member ensemble for SAI under CESM2 is inherently weaker than 20-member ensembles under CESM1. CESM1 has a shorter 30-year historical period from 1980 to 2009 which could not capture the interdecadal variability modes of the teleconnection patterns."

The following sentence has been added to the abstract:

"However, these findings are limited by the data available, especially for multi-decadal signals, with less than 100-year long simulations available for SAI."

- Finally, the figures, particularly figure 1, are complex and difficult to interpret. The authors should consider which elements are needed to make their argument and which might be consigned to supplementary material.

<u>Reply</u>: We have a completely new Fig. 1 (shown in the answer on a minor point below). We have revised all the figures in the light of specific comments as shown below, and the earlier replies. The old Fig. 1 is also presented in Supplementary Information, named "Figure S1". The former Fig. 1 is now Fig. 2 and changed as:

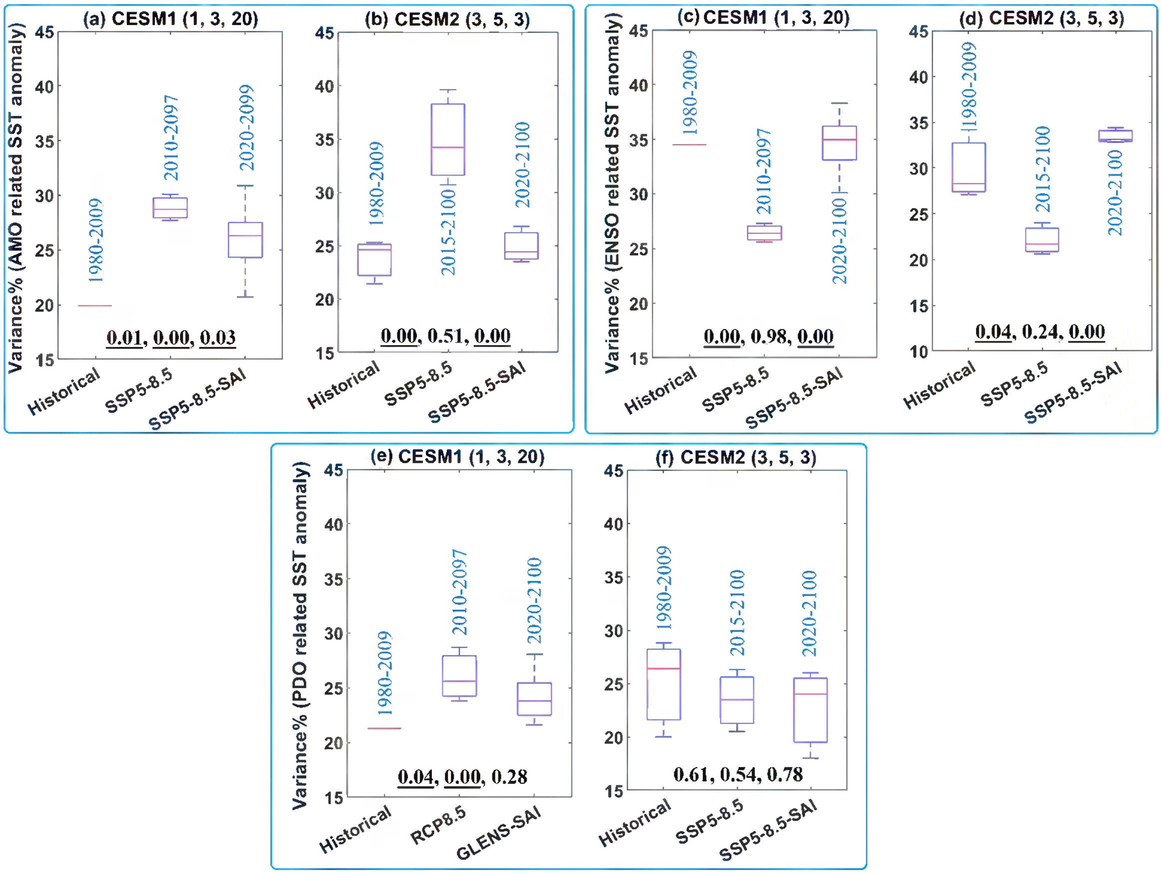

Figure 2. Box and whiskers plot of the variance in the leading EOFs, representing AMO, PDO, and ENSO, relative to the total variance of the SST fields: AMO across the North Atlantic (top-left panel); ENSO (top-right panel) global SST; and PDO across the North Pacific (bottom row). The values in blue on each column box show the period of the data for historical, greenhouse gas (i.e., RCP-8.5 and SSP5-8.5), and climate intervention (GLENS-SAI and SSP5-8.5-SAI) scenarios. The titles of each subplot refer to the CESM version and the number of ensembles used in the historical, greenhouse gas, and SAI scenarios, respectively. The median for each experiment is denoted by the red line, the upper (75th) and lower (25th) quartiles by the top and bottom of the box and ensemble limits by the whisker extents. The three values shown at bottom of each sub-plot refer to the p-values obtained from the statistical t-test between historical and global warming, historical and SAI, and global warming and SAI, respectively. Values underlined are significant (i.e., $p<0.05$).

Minor comments

- The authors could consider adding to Figs 2,3,4,5 a row showing differences under SAI, and an indication of significance of the magnitude this difference. Without such a row it is difficult to interpret the impact of SAI in these figures. The authors might also consider moving all these spatial figures to supplementary.

Reply: We plot the ensemble mean pattern and added maps for RCP8.5 minus historical, GLENS-SAI minus historical, and GLENS-SAI minus RCP8.5 where the stripping pattern reveal not statistically significant changes, based on p-values from t-test analysis. Accordingly, more explanations have been added to the main text of the paper.

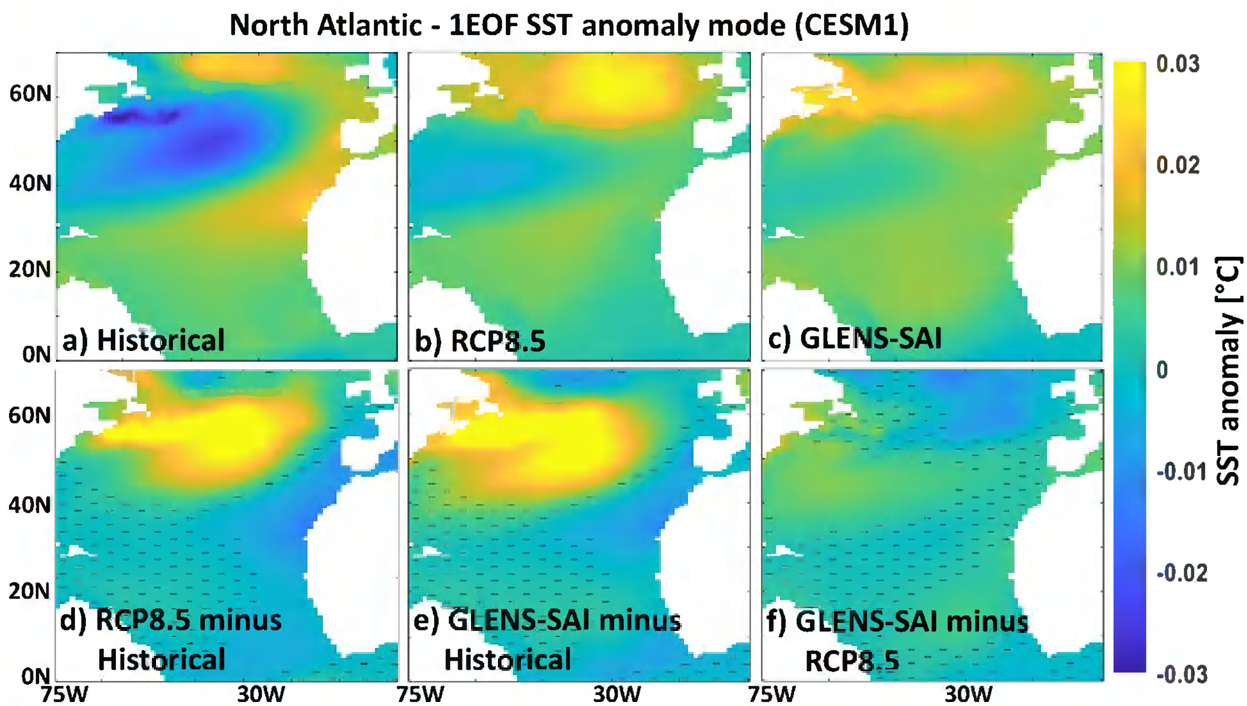

Figure 3. The first EOF (1EOF) patterns of SST anomaly across the North Atlantic relate to the AMO index simulated by CESM1 for the historical data (a) and the mean of the available ensemble members outputs under the RCP8.5 (b) and GLENS-SAI (c) scenarios. The maps at the bottom row show RCP8.5 minus historical (d), GLENS-SAI minus historical (e), and GLENS-SAI minus RCP8.5 (f) where the hatched patterns are not statistically significant (p>0.05), based on p-values from t-test analysis.

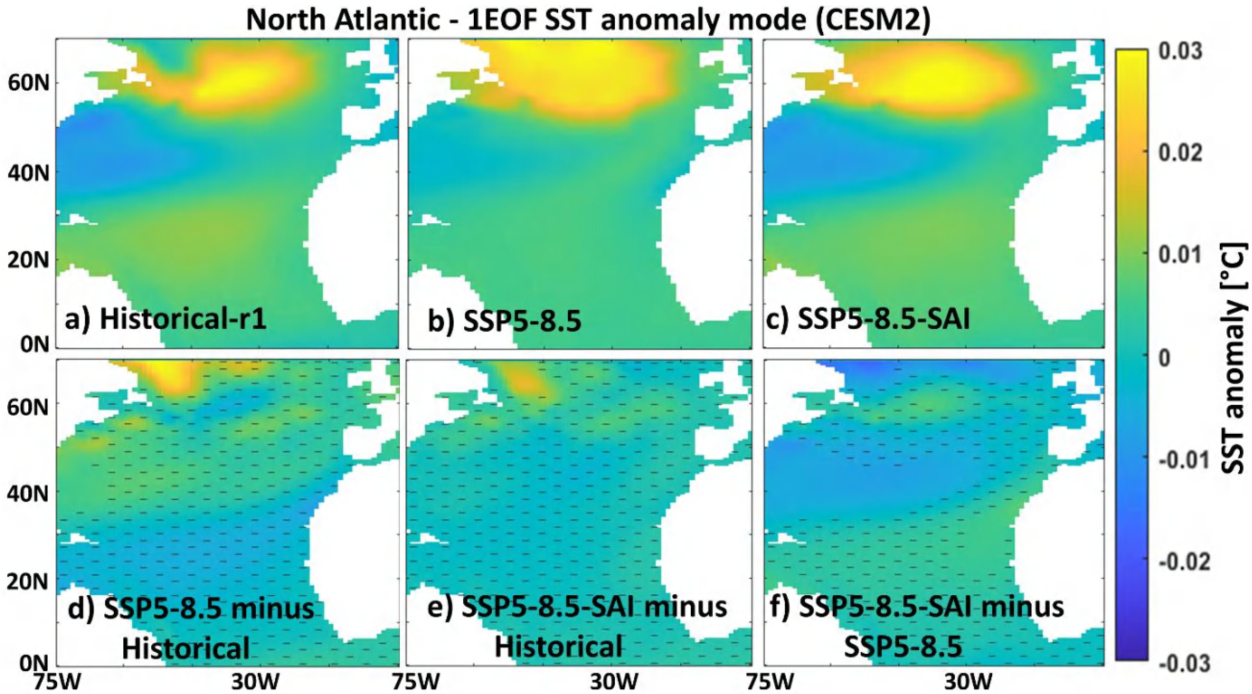

Figure 4. As in Fig. 3, but for CESM2 and SSP5-8.5.

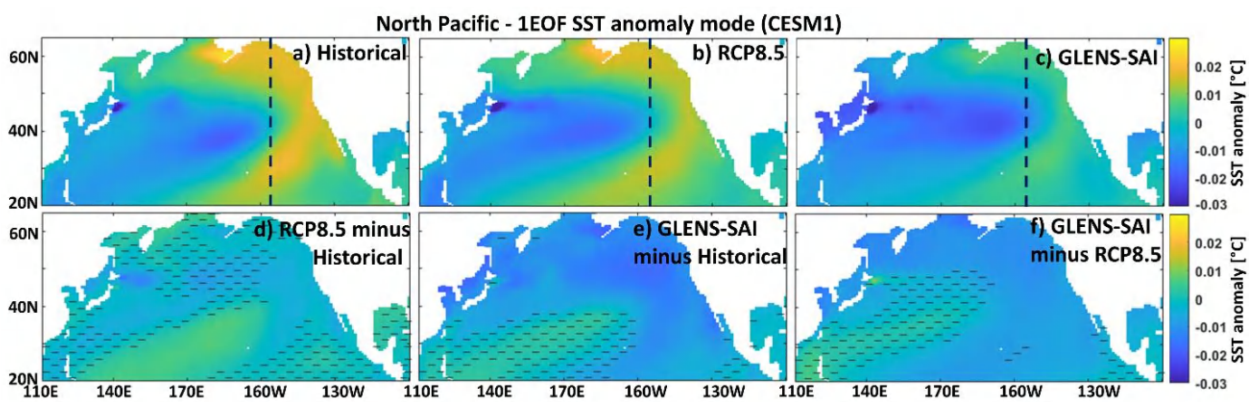

Figure 5. As Fig. 3 but across the North Pacific relate to the PDO index.

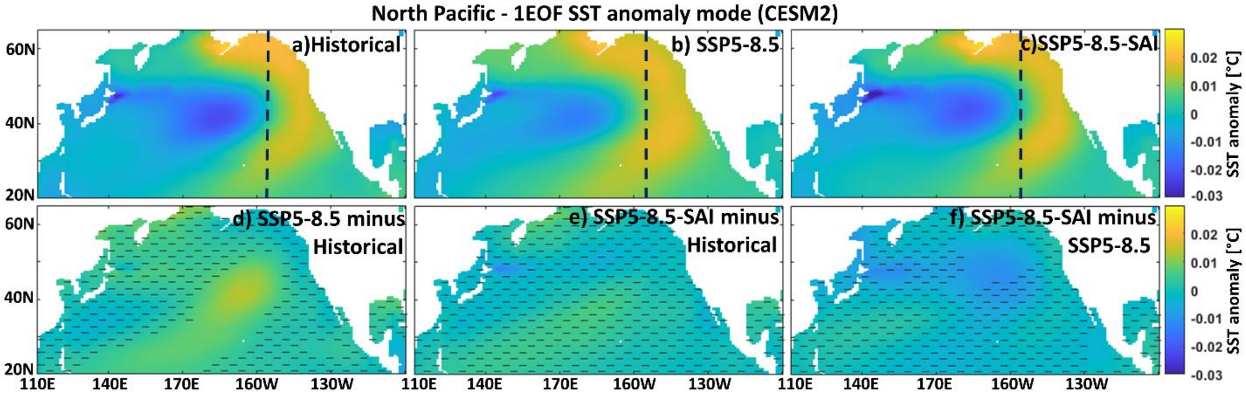

Figure 6. As in Fig. 5, but for CESM2.

- Figure 8 shows a 100-fold increase in NAO power in the high frequency end of the spectrum between the historical and the SSP5-8.5/SAI scenarios. This result is not discussed in the text but is very surprising. The authors should explain what is happening here, and address whether the finding casts doubt on the ability of CESM2 to capture NAO variability.

  Reply: We have re-checked all the steps and the codes from scratch, it was due to a small error in the code we wrote to compute the historical NAO from CESM2 PSL data. After correction, the power of the historical NAO in the new Fig. 8 is similar to the future simulations:

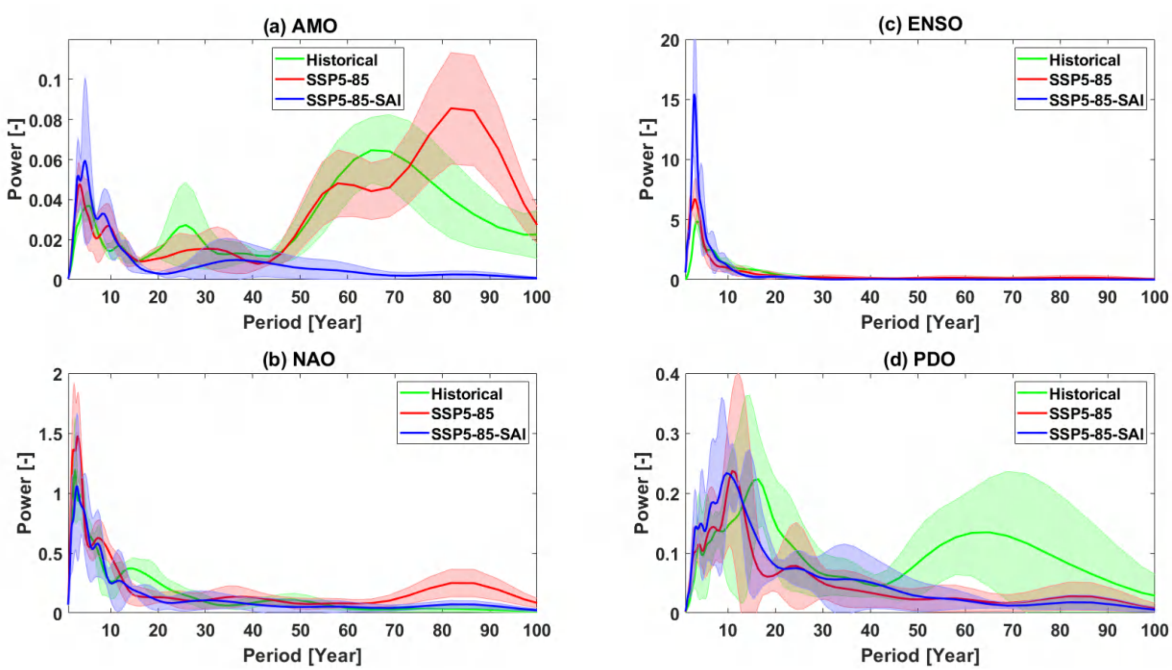

Figure 8. The CWT global power spectrums obtained for the indices of (a) AMO, (b) NAO, (c) ENSO, and (d) PDO under SSP5-8.5 and SSP5-8.5-SAI relative to the historical results based on CESM2 for the periods of 1850-2014. Shading in each curve shows the across-ensemble range.

- The authors might consider removing Figures 7 and 8a-d, since they are somewhat misleading in suggesting that the historical run differs from the other runs in the high period end of the spectrum when in fact it is simply too short to represent this part of the frequency space.

Reply: We have removed the old Figures 7 and 8a-d and their corresponding explanations from the paper.

- For the supplementary figures S1-S4, the authors should consider grouping these plots by index rather than by simulation, and showing all simulations vertically for each index so that a comparison can be made. I would also suggest adding at least one of these index timeseries figures to the main body of the paper.

Reply: We have shown all the indices from r1-r3 for the SSP5-8.5, and SAI scenarios in the new Fig. 1.

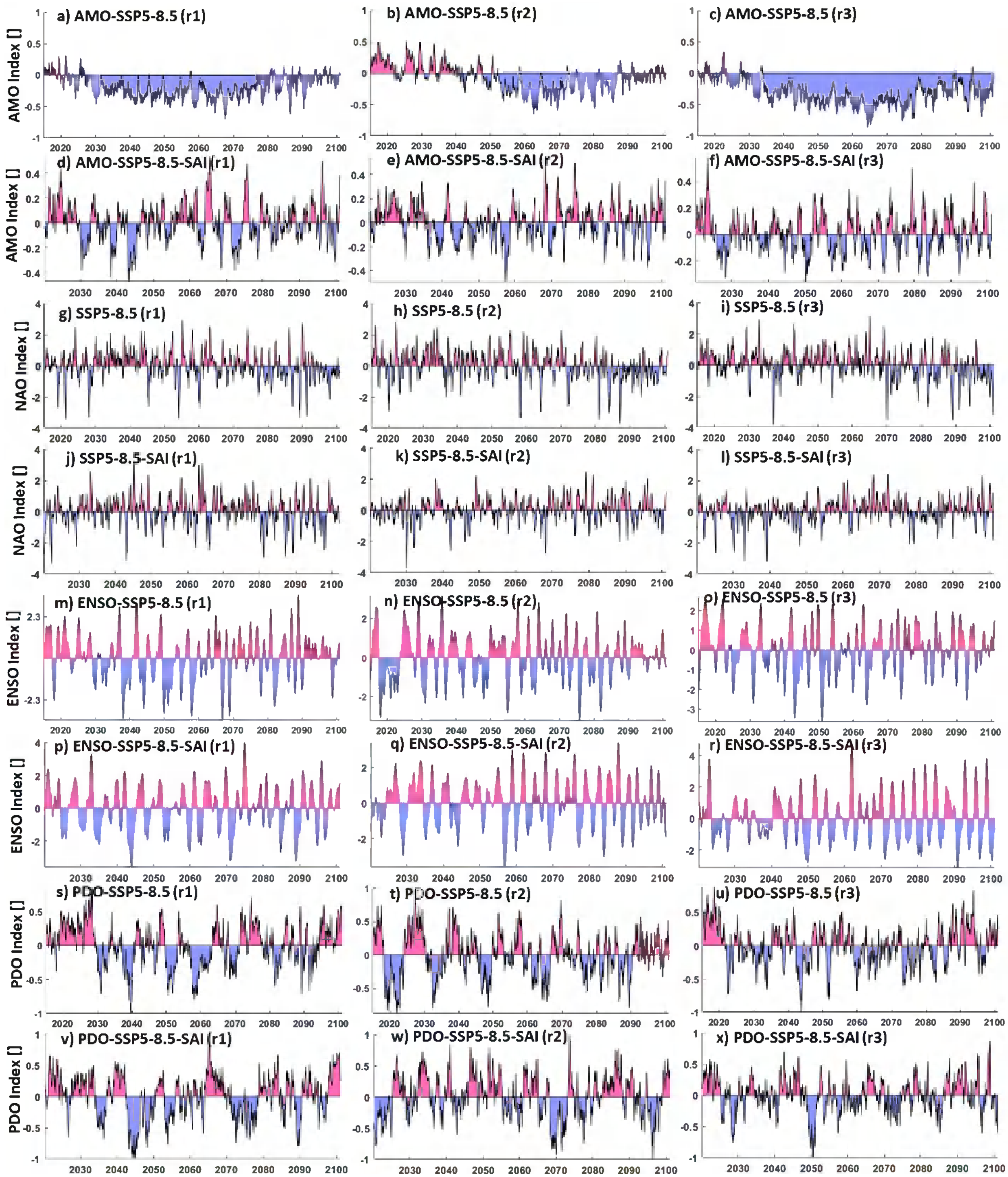

Figure 1. AMO (panels a-f), NAO (panels g-l), ENSO (i.e., NINO3.4, panels m-r), and PDO (panels s-x) indices obtained from ensemble members r1(left column), r2 (middle column), and r3 (right column) of the SSP5-8.5 (odd rows) and SSP5-8.5-SAI (even rows) scenarios.

Line 18 (and throughout): The authors might consider their use of the phrase “climate teleconnection patterns” to describe features such as the Atlantic Multidecadal Oscillation.

Reply: Implemented.