

Reply on RC1

Keith B. Rodgers et al.

Author comment on "Ubiquity of human-induced changes in climate variability" by Keith B. Rodgers et al., Earth Syst. Dynam. Discuss., <https://doi.org/10.5194/esd-2021-50-AC2>, 2021

We thank the reviewer for their careful reading of the manuscript. In our responses detailed below to the review, we have dedicated our attention wherever possible to the suggestion that we take the opportunity to provide an interpretation of our findings. To this end, we have emphasized in our responses where we plan to anchor the descriptions of our large ensemble behavior to mechanisms and model behaviors identified in previous studies.

Our detailed responses are below. For purposes of clarity, we have put the reviewer comments/questions in italicized text, and responded in plain text.

SPECIFIC COMMENTS

ll 33-37: I find a bit limiting the notion of fluctuations as "characterized by spectral variance peaks superimposed upon a broad noise background", as I think it does not entail the possibility that modes of spatio-temporal variability are in fact influenced by the "noise background" itself. Especially when one deals with processes that have clearly non-Gaussian PDFs, as in the case of this analysis, it is worth mentioning that at least multiplicative noise processes (and externally-driven changes therein) can alter the modes of variability through nonlinear interaction (e.g. Majda et al, 2009; Sardeshmukh and Sura, 2009; Sardeshmukh and Penland, 2015).

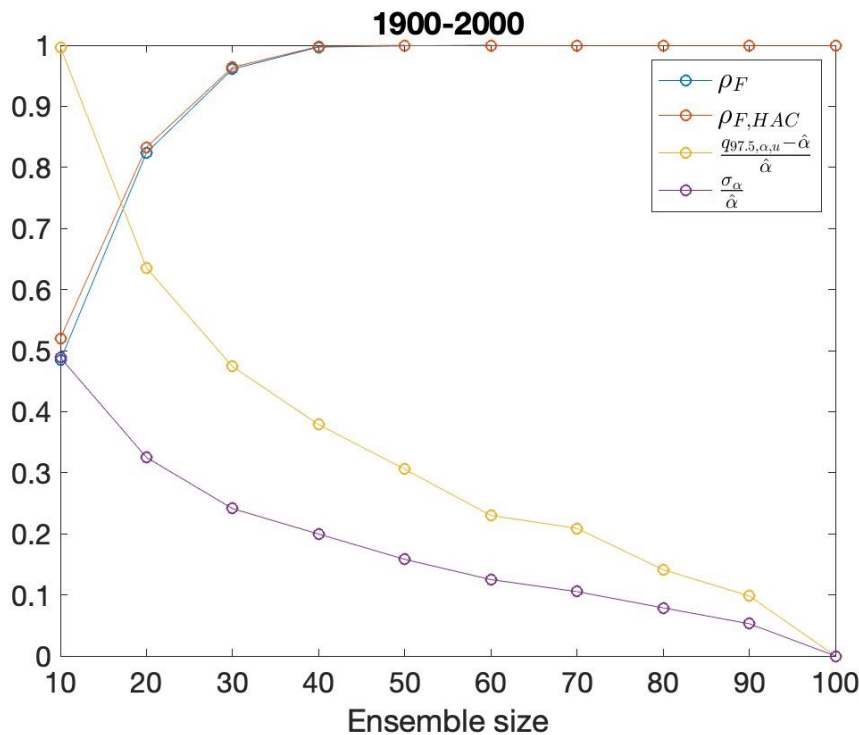
We thank the reviewer for raising this point. Indeed the role of multiplicative noise in affecting variability on a multitude of timescales is well established in the literature. The references suggested by the reviewer will be included in the revised draft, in addition to references to Levine and Jin (2010; JAS) and Jin et al. (2020, Simple ENSO Models in AGU monograph on El Niño in a changing climate). To our knowledge the Mueller paper is the first that describes the influence of multiplicative noise on second- and third-order cumulants and spectra in the context of the stochastic climate model. The other papers touch on the highlight the role of multiplicative noise in generating ENSO characteristics.

We will further revise the text by adding the following sentence: "The spectrum of observed regional-to-global climate fluctuations exhibits spectral variance peaks and a

broad noise background (Hasselmann, 1976; Franzke et al., 2020). Spectral peaks can emerge from a range of mechanisms, including astronomical forcings and internal climate instabilities, such as for ENSO. Moreover, these distinct features can be further influenced by climate processes acting on different timescales. Examples for this type of nonlinear “timescale interaction” are multiplicative (state-dependent) noise (Mueller, 1987; Majda et al., 2009; Sardeshmukh and Sura, 2009; Sardeshmukh and Penland, 2015; Jin et al., 2007; Levine et al., 2010; An et al., 2020; Jin et al., 2020) and combination model dynamics (Stuecker et al., 2015).

ll 62-64: while I find that mentioning Milinski et al. (2020) objective algorithm for the detection of the required LE size is appropriate, I think that with the algorithm being model-dependent, it should be acknowledge that their conclusions do not a priori apply here. Possibly a sampling over the pre-industrial simulation, using it to test the internal variability associated with ENSO, would hint at the number of members that is actually required (even though one would have to assume that the same holds when the SSP3-7.0 forcing is applied).

The point of the reviewer regarding the model-dependence of the Milinski et al. (2020) study is well-taken, and this question is the subject of an independent study with CESM2 led by one of our coauthors, Tamas Bodai. It is indeed not possible to project the necessary ensemble size in a model based on findings from another model configuration, and there are other issues that complicate the idea of prescribing a general rule ensemble sizes required for an experiment. As for the generic idea of an ensemble-size dependence of detectability and the accuracy of identifying forced changes, the following figure considers variance changes for Niño3 SST following the suggestion of the reviewer:



In the figure, rho denotes the detection rate, calculated with a bootstrap mean over 1e3 bootstrap samples, and F refers to the F-test for the slope, concerning whether it is

significantly non-zero at the 95% significance level. HAC refers to a technique that considers heteroscedasticity and auto-correlation – which are themselves in fact not detectable here, hence the agreement between ρ_F and $\rho_{F,HAC}$. Thus the 20th century changes in the model are already detectable with 20 ensemble members.

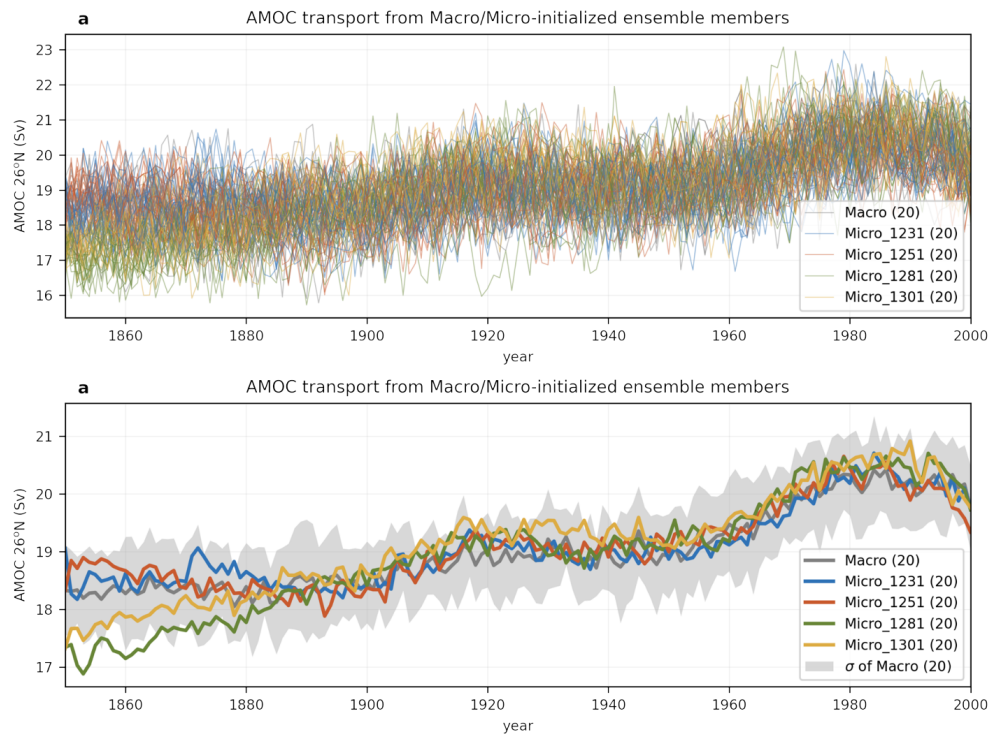
The study of Milinski (2020) is not concerned with detectability, but rather with the accuracy of large changes, and this is addressed by the yellow and purple lines in the figure. α is the temporal slope of the ensemble standard deviation, and the purple line indicates the best estimate from the 100 ensemble members. $q_{97.5}$ and σ are the 97.5th quantile (corresponding to the upper (u) bound of the 95% confidence interval) and the standard deviation over the bootstrap samples, respectively, with the relative errors/"variance" plotted.

ll. 106-107: the choice of the section of the pre-industrial run, where the model drift is particularly small, should be better justified. The internal variability of the model might be influenced by the presence (or absence) of such bias, and it would be relevant to assess how relevant this impact is.

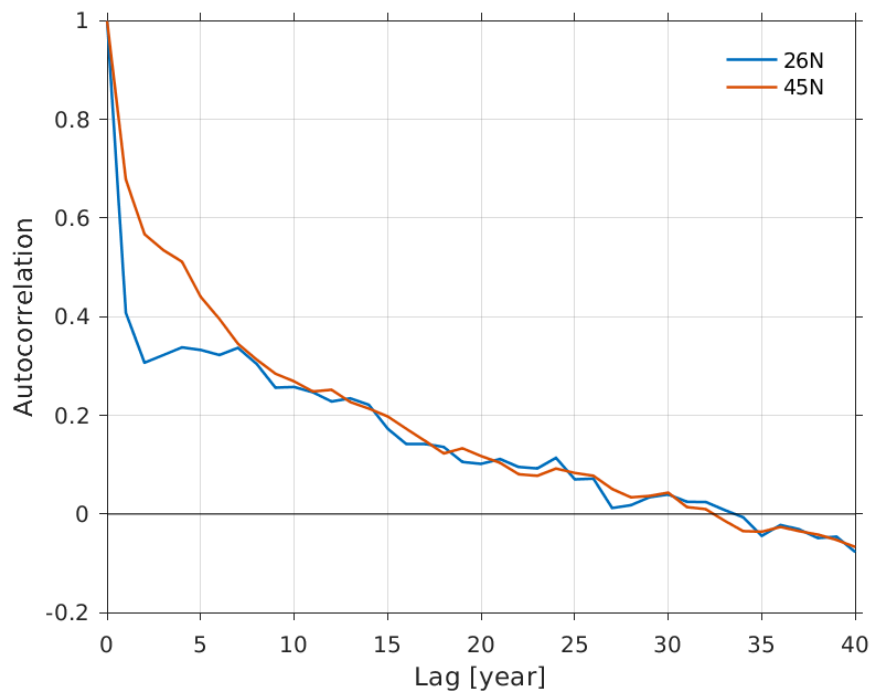
In the revised manuscript, we will provide a clearer description of model drift over the span of the initialization dates from the pre-industrial control run. In the CESM2 presentation paper of Danabasoglu et al. (2020), Fig 6. Showed the TOA energy imbalance for the pre-industrial run over years 1-1200, revealing minimal drift by year 1000. Additional extensive diagnostics by NCAR scientists identified minimal drift for the AMOC, and for upper ocean heat content over the North Atlantic and Southern Ocean over years 1000-1300.

ll 108-116: I am a bit puzzled by the choice of the initialization dates for the ensemble. 80 members are initialized with 4 initial dates (sampled according to the phase of the AMOC; maximum AMOC, minimum AMOC, ascending AMOC, descending AMOC), then slightly perturbing these initial conditions (20 members per date); for the additional 20 members, initial dates separated by 10 years were chosen. I find hardly justifiable that the members are to be considered as independent, and identically distributed, and that as such, conclusions can be drawn about ensemble mean moments of the distribution. I acknowledge that, as the authors state at ll. 122-123, "further quantitative exploration of the specific duration over which initial condition memory is retained is the subject of a separate ongoing study", but I see two issues in this choice of the initial dates: 1. Members chosen according to AMOC phase are not uncorrelated by construction; 2. When it comes to the internal variability of the ocean, it is quite unlikely that 10 years are a sufficient decorrelation time.

We were not sufficiently clear in our submitted manuscript about issues surrounding the initialization strategy, through a mix of macro-and micro-perturbations. It was not our intention to argue that the initialization procedure for CESM2-LE produces members that can be considered as independent, and we should have stated this more clearly, we apologize for the misunderstanding. Our analysis with internal variability is primarily focused on the post-1960 period, so more than a century after the 1850 initialization time for all members. In order to clear up any potential confusion, in addition to more detailed text clarifying the initialization procedure, we will provide in the revised manuscript both a quantification of the decorrelation timescale for the AMOC, as well as a timeseries figure showing the evolution of AMOC transport over the late 19th century for the 4 sets of micro-perturbation runs, illustrating the timescale over which initial condition memory is lost.



This two-panel figure shows the evolution of AMOC for (top panel) individual ensemble members over 1850-2000, and (bottom panel) ensemble means for the four micro-perturbation groupings as well as the ensemble mean of the macro-perturbations, all considered at 26.5°N. By the year 1900, the bottom panel indicates convergence in the ensemble mean groupings.



This is consistent with the autocorrelation for the (detrended) AMOC calculated using years 401-2000 from the pre-industrial control run (piControl), as shown for both 26N and 45N in the above figure. This is consistent with our interpretation that the analyses in Fig. 2 and Fig. 4 occur well beyond the time when initial condition memory is important.

ll 130-136: I do not think that enough evidence is here provided that the two ensembles with different biomass burning can be assumed as being (or not being) part of the same population. An assessment through statistical tests (e.g. Mann-Whitney?) would here support such an argument.

We fully agree with the reviewer that during the period of biomass burning perturbations (1990-2020, effectively) the full suite of 100 members should not be assumed to be part of the same population, but rather considered as two sets of 50 members. This is what motivated our Supplementary Fig. 2 in the submitted draft. There are two new manuscripts under development led by ICCP scientists (coauthors on this study) that deal explicitly with the impacts of biomass burning on the climate state. For the surface temperature, sea ice, and precipitation the response is only significant over the 1990-2020 interval of the biomass burning perturbation itself. This is the reason why we chose the intervals 1960-1989 and 2070-2099 for our emphasis on changes in variance, as the first of these is prior to the biomass burning perturbation, and the second is 50+ years after the perturbation.

Upon further reflection, we also recognize that we have not been sufficiently clear in the main text about which members are grouped with the CMIP6 and SMBB representation of biomass burning. To that end we will include as Supplementary two schematic figures that have already been prepared for our online description of the model runs:

<https://www.cesm.ucar.edu/projects/community-projects/LENS2/>

namely the two figures shown immediately above, in our Supplementary Materials to facilitate understanding of the model output organization. Additionally, we will mention more clearly in the revised text that the code base for the SMBB simulations incorporate corrections to minor bugs that were present in the first 50 ensemble members. This pertains to the SO₂, SO₄, and SOAG emission datasets. For SO₂ and SO₄, the “shipping” and “agriculture+solvents+wate” components of forcing were inadvertently switched during the historical-to-projection transition, or more specifically in 2015. The bug with SO₂ partitioning does not impact the results given that its components are summed up

before use. On the other hand, the issue with SO₄ datasets can impact the model state evolution because the particles sizes and numbers differ for the SO₄ components. The SOAG emissions are calculated from several hydrocarbons, and they were not recalculated after an earlier bug correction in covering units of the lumped species for the biomass burning emissions. This issue was corrected, and diagnostics indicate that there are not any significant changes in the model solutions from these particular corrections with aerosols. As a related point, we will also state more explicitly that a bug corrections were introduced for the 50 SMBB simulations that correct for sporadic large CO₂ uptake over land that occurred for the CMIP6 runs due to a negative flux of carbon, occurring at crop harvest time in a single time step. Although these large negative carbon flux component terms in autotrophic respiration are necessary for maintaining carbon balance, the large CO₂ spikes are not desirable. To avoid these spike, the associated CO₂ flux that occurred over a single time step for “dribbled” to the atmosphere over a time scale of approximately ½ year for the SMBB simulations. Our evaluations indicate that these bug fixes for carbon did not result in any climate-changing impacts for these modifications.

ll 175: out of curiosity, I was wondering why the authors chose to take into account the maximum transport at 40°N, instead of 26.5°N (which is often considered as an AMOC metric).

We agree with the reviewer that for consistency, the AMOC as represented in both Fig. 1 and Fig. S5 should be analyzed at 26.5°N. We will modify Fig. S5 accordingly for the revisions.

ll 201-202: as the authors refer here to variance and extremes, and their changes in future climate, it might be worth noticing that some promising results have been achieved with methods that synthesize several or all moments of the PDF, e.g. the Wasserstein distance (cfr. Ghil 2015; Robin et al., 2017; Vissio et al., 2020 for a climate model diagnostics application).

We thank the reviewer for pointing to these earlier publications, we will reference them in the revised manuscript.

ll 227: I do not have a clear idea for why the authors decided to retain the seasonal cycle in this context

We thank the reviewer for raising this point regarding the retention of the seasonal cycle in the wavelet analysis shown in Fig. 3, as we were not sufficiently clear about this in the submitted manuscript. Our reason for retaining the seasonal cycle stems from our interest in illustrating timescale interactions between ENSO and the seasonal cycle with the full power of large ensemble statistics. It is our hope that this will stimulate, as part of our presentation, further investigations of insights that are offered into frequency entrainment, among other questions that could arise. In our revised manuscript we will be clearer in making this point, and providing appropriate references in the context of explaining why we included the seasonal cycle.

The annual cycle and ENSO interact with each other in a complicated way, with the annual cycle itself being a forced mode (Xie, 1994). This interaction gives rise to combination models (Stuecker et al., 2015), frequency entrainment (Timmermann et al., 2007) and ENSO's phase-locking and seasonal variance modulation (Stein et al., 2014; Stein et al., 2010). Not only does the annual cycle in the equatorial Pacific influence the amplitude and phase of ENSO, but also vice versa. Due to this intricate coupling between these modes of variability, we have decided to retain the seasonal cycle in this context.

This information will be explicitly conveyed in the revised text, and we appreciate the reviewer for having raised it.

II. 246-247: this is one of a few sentences I found in the text, that justify my general comment above about the lack of interpretation. In particular, the authors mention a lead-lag relation between precipitation and SST seasonal maxima. The assessment of these relations are challenging in the context of climate models (Lembo et al., 2017), together with their interpretation (cfr. Su et al. 2005 for this specific context) and the authors might want to discuss what these mean in terms of dynamics of the systems.

We agree with the reviewer that there is an opportunity here to reference published literature that presents mechanistic interpretations of the behavior we have highlighted for the CESM2-LE. For the specific issue raised here of maximum precipitation leading maximum temperature over the Niño3.4 region on seasonal timescales (red dots in Fig. 3c and 3d), current scientific understanding maintains that precipitation is largely driven by meridional SST gradients, and is thereby not directly tied in its phasing to local SST. In other words, moisture convergence is in part determined by non-local SST conditions. We will appropriately reference the study of Xie (1996), Xie et al. (2010), Williams and Patricola (2018), and Stuecker et al. (2020) on this topic in the revised text.

I. 276: This is in part already known. Several studies (e.g. Screen 2014, Chen et al. 2016, Haugen et al., 2018) have shown evidence of a relationship between Arctic amplification and reduced temperature variance over the mid- and high-latitudes of the Northern Hemisphere, and an interpretation of this has been given from a dynamical point of view (cfr. Sun et al., 2015; Schneider et al., 2015), involving the role of precipitation.

We thank the reviewer for suggesting that appropriate references be given for describing changes in variance in temperature over the mid-to-high latitudes of the Northern Hemisphere due to polar amplification. We will reference these studies in the revised version of the manuscript, as well as the studies of Holmes et al. (2016; J. Climate) and Screen et al. (2015; BAMS).

II. 322-323: same as in my comment to I. 276. I am not surprised that the authors find a reduction in the NEP inter-annual variability, as this is linked to the variability of near-surface temperature. The link has been discussed in previous works (e.g. Yao et al., 2021), and I believe it should be taken into account here.

It seems that the reviewer may in fact be confused by the text in our manuscript. We specifically chose not to address interannual variability in NEP. Rather our focus was on interannual variability in phenology in the lower panel of Fig. 5, as this is the behavior that has to our knowledge not been previously described in published literature. In our revisions, we will move the interpretation of the cause of the forced trend earlier in the paragraph, where we discuss the time of emergence of the trend.

In order to clarify, we will modify text at the end of our discussion of phenology to say: "Our analysis indicates that for NEP aggregated over this region the phenological shift as a decadal trend becomes emergent relative to estimates of the natural variability already within the first decades of the 21st century, a trend that is broadly consistent with observations (Zhu et al., 2016; Myers-Smith et al., 2020). The forced changes in growing season length are mostly attributable to changes in the mean temperature (Lawrence et al., 2019; Lombardozzi et al., 2020). Internal variability in the date of the onset of the growing season decreases by 35% over the course of the simulations and decreases by 18% for the date of the end of the growing season (Fig. 5, lower panel)."

The statements pertaining to attribution (expansion of growing season length being attributed to temperature) will be moved to the beginning of the same paragraph, so that the attribution is more of an emphasis in the text.

II. 328-330: see my comment at II. 246-247. I think the authors should comment on this

finding and on how this can be interpreted.

The paragraph pointed to is introducing the climate change impacts on the mean state and variability. The reviewer is asking us to interpret the result related to changes in variability of peak and trough NEP amplitude (mentioned in lines 328-330, changes in variability along the y-axis of Fig. 5), but we don't believe that this is where the main story lies here. Instead, we focus on the discussion of the temporal trends and changes in the seasonal variability related to spring-green up (x-axis of Fig. 5). Specifically, we focus on changes in the mean state related to forced changes in phenology, notably the earlier initiation of the growing season in the Northern Hemisphere spring, where ecological functions may be disrupted. Under future scenarios we also see reduced variability in the first zero-crossing of NEP, which we attribute to the combined effects of warming and the timing of snowmelt. A subset of the coauthors of this manuscript are pursuing this question of drivers of phenology as an offshoot project that will complement the presentation paper analysis.

ll. 350-351: this is not a new achievement, it has been long known (see e.g. Palmer, 1993; Corti et al., 1999) that climate change projects on modes of variability in several ways.

We fully agree with the reviewer that the concept of anthropogenic changes can project onto modes of variability has been known for some time, and we will reference the studies of Palmer (1993) and Corti (1999) in our revisions. In order to better clarify this point, we will also state more clearly that the Large Ensembles provide a means to explore this in a statistically sound way in full complexity climate models that was not yet possible in the 1990s.

ll. 360: I wonder if the authors are able to comment on how significant these findings obtained with CESM2-LE are, in relation with other Large ENsemble exercises. described in Maher et al. (2021).

The reviewer refers to the multi-model Large Ensemble intercomparison study of Maher et al. (2021), and by implication the growing number of studies that make use of the multi-model Large Ensemble Archive presented in the study of Deser et al. (2020). From the onset of our project with the CESM2-LE, our intention has been to have our simulations be available for such multi-model studies, and to that end we specifically chose the historical/SSP3-7.0 pathway recommended in the CMIP6 protocols (ScenarioMIP) study of O'Neill et al. (2016). We have opted here to not engage in an intercomparison exercise, as this is beyond the scope of this initial presentation of the CESM2-LE itself, but to reiterate we have made every effort to facilitate such inter-comparison studies by interested parties in the future. We have also made available (<https://climatedata.ibs.re.kr/data/cesm2-lens/lens-diagnostics>) the results of the Climate Variability Diagnostics Package for Large Ensembles (CVDP-LE) of Philips et al. (2020) for diagnostics over a broad suite of variables that can equivalently be run for other large ensembles, as a means to facilitate studies that seek to understand model differences. We will include a link to this with an explanation in the revised version of the manuscript.

ll. 364-367: the lack of interpretation of the findings is evident here. I don't think that the take-home message is that the Earth system is "far more sensitive in its statistical characteristics to anthropogenic forcing than previously recognized". There is actually a literature on assessing changes in higher order moments of several aspects of climate variability, often using Large Ensemble exercises, e.g. Swain et al., 2018, for regional precipitation, Tamarin-Brodsky et al., 2020, for NH temperature variability, among others. The authors might compare their results with others, in order to explain how the sensitivity of statistical characteristics was less recognized before. As mentioned above, some of the findings, taken one by one, are confirming, or possibly expanding, what was

already somehow known from previous works. The manuscript might be significantly improved, if the authors would at least qualitatively discuss what drives, and what is the relation between e.g. changes in frequency and phasing of ENSO w.r.t. SST and precipitation, cross-ensemble SD for temperature and precipitation, change in ENSO's remote correlation with regional mean temperatures and precipitation over some regions, just to mention a few features that might be interpreted in the light of changes occurring in the general circulation.

Yes, we agree with the reviewer that the sentence towards the conclusions of the manuscript that "the Earth system is far more sensitive in its statistical characteristics to anthropogenic forcing than previously recognized" would be better rephrased as "we have provided support with new examples and new global emphasis that the Earth system is sensitive in its statistical characteristics to anthropogenic forcing, thereby building on previous studies."

With regard to the questions posed by the reviewer with regard to Fig. 4, namely ENSO teleconnections, we will also reference properly the AGU monograph published in 2020 entitled "El Niño Southern Oscillation in a Changing Climate", in particular the chapter by Taschetto et al. entitled "ENSO Atmospheric Teleconnections". This synthesis reference appropriately addresses the challenges for understanding how ENSO teleconnections can change, including the relative role of local diabatic forcing and modulations of ENSO for understanding regional responses.

TECHNICAL CORRECTIONS

ll. 9: replace "runs" with "run"

We will correct this error in the revised manuscript.

ll. 150: replace "weaking" with "weakening"

This will also be corrected, following the suggestion of the reviewer.

ll. 209: replace "associate" with "association"

This will also be corrected.

ll. 256: replace "n" with "in"

This will also be corrected.