The manuscript features an evaluation of several key components of the climate system, focusing on statistics of their variability and how they change as a consequence of anthropogenic climate change. The analysis makes use of the largest (among the recently performed) Large Ensemble (LE) experiment, carried out with CESM2 model. In order to assess changes in the statistics, historical and SSP3-7.0 scenarios are considered, and decadal averages are performed. The authors find that the signature of climate change is apparent, not only in the mean state change, but also in the variance, in the occurrence of extreme events, in the amplitude and frequency of certain periodic oscillations, and in some aspects of co-variability of selected quantities.

Overall, I think that the manuscript is scientifically sound, the methodology is reasonably correct in its implementation, consistently with the aim of exploiting the opportunity given by the huge CESM2-LE dataset to perform an in-depth analysis of climate variability from a global-scale point of view. Nevertheless, I think that the authors miss the chance to provide an interpretation of their findings. As a result, the manuscript is characterized by a collection of outputs loosely connected with each other. Secondly, a few methodology aspects deserve more careful consideration, such as bringing together runs with different treatment of biomass burning fluxes, selecting the initialisation points according to the phase of the AMOC, or considering 10-year period as a sufficiently long decorrelation timestep for the sampling of initial conditions. Finally, I think that the manuscript would benefit by better referencing the available literature, particularly on the impact anthropogenic climate change has on variance and extremes.

Once these points, that are specifically addressed in the comments below, are taken into account, I think that the manuscript could be accepted for publication.
SPECIFIC COMMENTS

II. 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 entails the possibility that modes of spatio-temporal variability are actually 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 that 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);

II. 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, being the algorithm model dependent, it shall be acknowledged 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).

II. 106-107: the choice of the section of the pre-industrial run, where the model drift is particularly small, shall 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;

II. 108-116: I am a bit puzzled by the choice of the initialisation dates for the ensemble. 80 members are initialised 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;

II. 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;

II. 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 metrics);

II. 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 models diagnostics application);

I. 227: I do not have clear why the authors decided to retain the seasonal cycle in this context;

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 (e.g. 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 system;

I. 276: This is in part already known. Several studies (e.g. Screen 2014, Chen et al. 2015, Haugen et al. 2018) have evidenced the relation 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;

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 shall be taken into account here;

II. 328-330: see my comment at II. 246-247. I think the authors shall comment on this finding and on how this can be interpreted;

II. 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;

I. 360: I wonder if the authors are able to comment on how significant these findings obtained with CESM2-LE are, in relation with the other Large Ensemble exercises described in Maher et al. 2021;
I. 364-367: the lack of interpretation of the findings is here evident. 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 wrt. SSTs and precipitations, cross-ensemble SD for temperature and precipitation, changes 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 to the general circulation.

TECHNICAL CORRECTIONS

- l. 9: replace "runs" with "run";

- l. 150: replace "weaking" with "weakening";

- l. 209: replace "associate" with "association";

- l. 256: replace "n" with "in";

REFERENCES

midlatitude near-surface temperature and tropospheric baroclinicity: reanalysis data and AOGCM simulations, Earth Syst. Dynam., 8, 295–312