

Earth Syst. Dynam. Discuss., author comment AC3
<https://doi.org/10.5194/esd-2022-26-AC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on CC1

Nicola Maher et al.

Author comment on "The future of the El Niño-Southern Oscillation: Using large ensembles to illuminate time-varying responses and inter-model differences" by Nicola Maher et al., Earth Syst. Dynam. Discuss., <https://doi.org/10.5194/esd-2022-26-AC3>, 2022

Review of The future of the El Niño-Southern Oscillation: Using large ensembles to illuminate time-varying responses and inter-model differences

by Maher et al.

The manuscript by Maher et al. seeks to diagnose changes in ENSO in 14 single model large ensembles (so-called SMILEs). The manuscript builds upon a body of work that is often based on single members, individual models, or idealized models and so it represents an advance, particularly at resolving the decadal varying aspects of forced changes in variance - which the work shows can be important. The manuscript is clearly written, is explicit about its objectives, findings, and reasoning, and includes figures that are well-designed and clear. There is sufficient new material here to justify publication. Some aspects are frustrating - such as in cases where the robust take-home message seems to be that there is no robust take home message. Though basic questions go unanswered concerning the origins of inter-model contrast and mechanisms of change, the broader community has also struggled to answer these questions and so this work is not unique in this regard. That said I do have some minor suggestions for improvement. This includes the general suggestion that multi-model means not be used for many of the metrics because of the disproportionate variance in some models that is swamping out the means. Rather I think medians make more sense since the broader goal is to make generalizations about model behavior, which implicitly seeks to screen out outliers. I have various other relatively minor suggestions listed below but otherwise view the manuscript as suitable for publication.

Thank you for taking the time to review this paper and for your positive comments and helpful suggestions.

While we agree that the median would make more sense for broader generalizations we choose to use the multi ensemble mean (MEM) for easy comparison with previous work that uses multi-model means (MMM) from CMIP. Here, the main aim is to show the diversity of forced responses that go into such a MMM and demonstrate that these diverse individual model responses average out in the MEM to show something that looks very similar to the MMM from previous studies. As such we choose to use the MEM rather than

medians in this study. We will add a sentence on this point in the revised manuscript to make the use of MEM clear to the reader.

55: There have been various studies that show the improvement of ENSO simulation across CMIP generations. These seem useful to reference in this paragraph to provide context on the numerous SMILEs.

We will add this to the revised version.

119: I appreciate that there can be useful examples of models that suggest that no simple relationship between the present day and future exists but perhaps a more thorough exploration across all models and metrics considered here would be more convincing?

The aim of this manuscript is to provide the community with a detailed overview of how each model behaves and demonstrate the time-dependence of the ENSO response. We agree this is interesting, but is out of scope of the manuscript and would be a great follow up study on it's own.

126: It is worth calling out the very large differences between SSP370 and RCP85/SSP585 in regarding to early 21st sulfate aerosols and potential consequences for the evolution of ENSO.

Thanks for the comment – this will be added to the revised manuscript.

Figure 1: There is substantial white space in the 3x5 layout of the figure. I recommend changing to 4x4 and ensuring there is little white space as the figure is somewhat inefficient and difficult to read as is. Also it looks as though the vertical extent needs to be expanded as some lines go out of range. Note also that there is substantial noise in many of the time series. I suggest applying a smoother except where the variability is not irrelevant noise. That said, various models seem to have abrupt responses to forcing, such as volcanic eruptions and perhaps even biomass effects (in MIROC-ES2L), that is too abrupt to be explained by warming alone. The authors seem not to find this worthy of discussion? I recommend addressing it.

Thanks, once the Figure is formatted into the text the white space should not be a issue. Keeping the time-series as is acts to demonstrate the inherent variability due to different ensemble sizes. Given the ensemble sizes are different this is helpful to give an idea as to how noisy the results we have are. We will check the vertical extent in the revised version.

We agree the abrupt changes are interesting, however they are difficult to diagnose without a thorough analysis of each models individual forcing differences. We will add a text highlighting these differences and speculating on why they occur in the revised version.

Figure 3: The multi-ensemble mean seems to be dominated by EC-Earth3. Would a multi-ensemble median perhaps be more appropriate?

See comment at the beginning of the response.

Figure 4: Fonts are too small - minimum font including axes should be on par with main text. Monthly stddev lines should be make thicker.

This will be changed in the revised version.

The authors don't provide any hypotheses for the seasonality of the change in variance? Do none exist? Again the MIROC-ES2L increase at 2000 is quite notable. Is there no explanation for why this may occur? Other models show periodic changes in variability in the future. Is this just noise? Does it suggest that again some additional smoothing is needed to deal with some small ensemble sizes using monthly data? If one looks at CESM2 there are again suggestions of periodicity? What might drive this? I suggest reducing the range of the color bar to make colors in the figure more visible.

We will reduce the colorbar range. In terms of the seasonality we cannot answer this without a full feedback analysis, which is out of the scope of this study. We will add the following text to the manuscript.

Determining the dynamical cause for the increased ENSO seasonal synchronization in most of the models will require a detailed ENSO feedback analysis (e.g., Chen & Jin 2022), including assessing potential future changes in the "southward wind shift" mechanism (e.g., McGregor et al. 2012, Stuecker et al. 2013).

Figure 5: Perhaps put whiskers on Fig 5 corresponding to the 2 standard error range, which seems to increases in the latter half of the year? Might multi-model median be more appropriate due to sensitivity of means to a single model?

We will add another line for the median in this plot. Rather than add a error range we choose to show each individual model in light grey to illustrate what goes into the multi-model mean.

Why are 99 members of CESM2 used for some figures and only 50 used for others? This is not discussed at all and contradicts with Table 1. I'd review all plots and ensure that the # of members used is consistent with Table 1.

Thanks for pointing this out - we will make sure all 99 are used where there we only 50 in the revised version.

Figure 6: I again question the use of multi model mean rather than median given the outsized influence of some models (CSIRO).

See comment at the beginning of the response.

Figure 8: Since the abscissa is not symmetric I recommend that a vertical line at 0 be shown to avoid confusion. As is, the panels and particularly the top, are misleading.

This will be added into the revised version.

Figure 8: Doesn't the fact that obs have become more La Niña like suggest that we should have seen a reduction in variance?

If you look at individual members you can see that while it is more likely with a La-Niña like change that there is a reduction in variance, not all members do this - showing the

inherent noise in the system. The addition of 0 lines as suggested above will make this clearer.

Figure 9: There should probably two more sets of arrows at the top of the plot saying La Niña-like and El Niño-like such that they are on each plot.

We will add these assuming they don't make the plot noisier and harder to read.

226: Why not also infer changes from the CESM1-SF (and now CESM2-SF if you do add CESM1 to the figure)?

Using the difference between the all and the all-but forcing used in CESM1-SF does not necessarily give the same result as a single forcing experiment. For consistency we only use single-forcing not all-but forcing experiments. CESM2-SF was not available when the workshop where the data analysis was done for this publication was held.

Figure 10: reference to 'all models' is a bit confusing given there are only 2 models. perhaps state "both models" or better yet "both ensembles"? or one could include ALL CMIP6 DAMIP simulations to provide context for canesm5 and miroct6. Also I wonder why not show these in a similar fashion to Fig 9?

We will update the text to both models. We do not show these the same way as Figure 9 as the periods we are looking at are much shorter and the single forcing ensemble sizes are inconsistent within the same model. Some of the small ensemble sizes made it difficult to understand the ensemble mean – this is why we show each member as well as the mean in Figure 10.

255: The discussions based on MEM should be reconsidered in the context of the median I think, particularly when making categorical statements of models overall (since the mean is strongly weighted by only a few models with large variance and is therefore not representative).

See response at the beginning of this document.

I recommend that the paper include a discussion of possible mechanisms and paths forward for exploring.

It is unfortunate that there is so little consistency across models on many aspects. The reader is left to wonder a bit on what is the robust finding here relevant to nature, aside from little being consistent across models? Have the authors looked for connections between some of the metrics being shown across models (e.g. seasonality of mean state changes and ENSO variance)? Have the authors examined what systematic differences exist for models that do best in some metrics packages (e.g. CVDP) in the present day (if so it would be good to mention, if not it would be good to do)? or for which changes in variance in observations fall within the ensemble spread, though this may be a weak constraint but would still be good to mention.

The aim of our paper is to illuminate how ENSO behaves in all available large ensembles. Here, we can look at ENSO evolution over time due to the use of large ensembles and truly identify how each model behaves under a strong warming scenario. We agree that understanding why the models behave differently and trying to better understand consistency between models is an important question. However, this is out of the scope of our study, which already includes 10 Figures in the main text and 7 in the Supplementary. We hope this work will inspire others to look further into these new datasets. Our final point in the manuscript aims to highlight this point and hopes to inspire further studies that look into this question in more detail.

"This highlights the need for further research on the mechanisms of inter-model differences in ENSO projections. There is a rich diversity of future ENSO changes projected by climate models and more work is needed to understand which aspects of these projections are robust."

We will add the following at the end of this discussion "are robust, why the models differ and how they relate to the world as observed."

304: Projections of nearly every climate quantity are nonlinear in time (since radiative forcing is nonlinear). Is there really an expectation of linearity? and if not is this really a significant result?

Why there is not an expectation of linearity most previous studies have not been able to look at the time-dependent response to assess whether there is linearity. For example most previous studies have considered a future time period (e.g. 2000-2099) compared to a historical period (e.g. 1850-1950).

308: As stated earlier I'd avoid the MEM(ean) and use the median.

See comment at beginning of document.

334: last sentence needs a period.

Thanks this will be added.

None of the panels in the paper are labeled (e.g. A, B, C, ...). I think they should be for easier reference in this and other manuscripts that may cite specific figure panels from this work.

We choose not to add these as they will make the Figures overly busy. Other can still reference the individual Figures and it should be easy to identify individual models as they are named in the titles.