Review of “Evaluating Uncertainty in Aerosol Forcing of Tropical Precipitation Shifts” by Peace et al.

General evaluation:

This research performs a large size of ensembles by perturbing 47 (or 49?) physical parameters. 13 members are selected to estimate the uncertainty of aerosol effective radiative forcing (ERF) and to investigate its relationship with ITCZ position shifts. The effects of aerosol forcing to the ITCZ positon have been explained by an energetic framework. This research provides a different result that does not support the theory previously suggested. However, this conclusion is not robust because the uncertainty of internal variability is very likely not well estimated due to small size of members (only 4 initial condition runs). In addition, some of the explanations and discussions about the PPE results are questionable. The description about experiment design is not that easy to follow. Therefore, I suggest a major revision to this manuscript.

Major comments:

1. The description about experiment design is confusing. I have to read Sections 2 and 3 back and forth for several times, and try to guess the design. I think some more detailed explanations will help a lot. Some of the tables and figures in Supplement can be indicated more specifically in the main text to help readers follow the design.

   To my best understanding, 47 model physical parameters are perturbed. It looks like Table S1 is the list of 47 model parameters for PPE mentioned in line 102 in the main text, but Table S1 has 49 parameters instead of 47. It may be helpful to indicate Table S1 in the main text (if my guess is right) or add a paragraph to explain Table S1 in the Supplement to help readers follow the experiment design.

   Then 13 ensemble members (model variants) are selected based on their performance (i.e. diversity as mentioned in line 111). Would you please describe in more details about what is “model variant”? I guess model variant means a simulation with a set of perturbed physical parameter values that indicate the location of the model variant in the “parameter space”. If my understanding is right, it is suggested to somehow rewrite this part. The concept is very abstract to me and it took me quite a long time to figure this out.

2. In Section 2, it may be helpful to explain the concept of large ensemble simulations (or add some references). For example, how does the large set of simulations helps interpolate the “signal (the uncertainty in aerosol ERF?)” and

"noise (internal variability)". This approach is relatively new, and may not be well known to all readers at present time.

3. About the result discussion in lines 401-403, you mentioned that the relationship might be masked by perturbations to physical atmosphere parameters. The mechanisms mentioned here are meaningful *only if* the uncertainty produced by perturbing the parameter is significantly different from internal variability. I think add some discussion about the signal-to-noise analysis may help clarify the concept. I understand the member size is small and the internal variability seems not well estimated in this research (only 4 initial condition runs). Therefore, this is my main concern for the discussion about PPE.

4. The relationship between inter-hemispheric aerosol ERF and ITCZ position is bad. I think the bad relationship is probably not due to the mask effects from physical atmosphere parameters. It may be because the aerosol ERF is not the dominant factor in nature that influences inter-hemispheric energy flux (which drives Hadley circulation and influences ITCZ position). The simulations are driven by historical emission (including GHGs and anthropogenic aerosols) and fork into four different future scenarios. The aerosol ERF only accounts for a small part of the energy change, compared to GHGs. It may be easier to see the effects of aerosol ERF if the effects of GHGs can be excluded (see Wang et al. 2019).

   Wang et al. (2019) Climate effects of anthropogenic aerosol forcing on tropical precipitation and circulations. *J. of Climate*. DOI: https://doi.org/10.1175/JCLI-D-18-0641.1

   Another reason is the ITCZ position change belongs to a slow precipitation response (decades) through processes involving surface-atmosphere interactions. The precipitation change is shown to have small response to fast radiative processes. The details can be found in Myhre et al. 2017. The ITCZ position is mostly related to inter-hemispheric energy flux (follow Frierson's energetic framework). This is why ITCZ position is better correlated with inter-hemispheric total forcing, and worse correlated with aerosol ERF (plus, your aerosol ERF may include large internal variability). The correlation between inter-hemispheric implied total forcing and ITCZ position is not very high, but good and stable (Table 1), which also supports that this is a robust factor.

   Myhre et al. (2017) A Precipitation Driver and Response Model Intercomparison Project—Protocol and Preliminary Results, *BAMS*, DOI: https://doi.org/10.1175/BAMS-D-16-0019.1

This may also explain why the analysis using additional inter-hemispheric variables (Table 2) does not provide supporting results because all of them are fast radiative response (a few days to a few years).

5. The inter-hemispheric surface temperature also has better correlation with ITCZ position than aerosol ERF, but the relationship is very sensitive to the time period (Table 1). This is probably because the slow precipitation response must involve slow surface-atmosphere interactions/adjustments, but the processes are interrupted by volcano eruptions for a few years. The volcano eruption is a very strong impulse in aerosol amounts and can causes significant temperature change in a relatively short time period. Therefore, you see stronger correlation when the eruptions are included in the time series. Is it possible to exclude years influenced by volcano in this analysis?

Minor comments:

1. The HadGEM3-GC3.1 is perturbed by small initial conditions to estimate internal variability. Four members seems to be a very small set of ensembles. It might be better to increase the number of ensembles because the uncertainty of PPE needs to be compared with the internal variability. Or maybe you can use other methods to exclude internal variability (also see Wang et al. 2019).

2. L.165, I am not fully understanding the method quantifying ERF here. What is the purpose/meaning of "plus 1860 and 1975"?

3. Line 348, I don't follow this sentence. What is "a large fraction of the trend…"? Can I find this information in a figure or a table?

4. You may want to do a plot the same as figure 4, but for inter-hemispheric implied total forcing, because surface temperature is too sensitive to the time period selected (as shown in historical analysis, Table 1) and may not be the best choice of variable for analyzing future projection.

Editorial suggestions:

1. Fig.S3, in the figure caption, "Historical emissions are shown in black, RCP8.5 in red and RCP2.6 in blue." is misplaced and can be eliminated.
2. Fig.S11, the title of subfigure (bottom) is "Atlantic", but it is "Pacific" in figure caption. Please check which one is correct.
3. Add detailed indications of sub-figure in the main text may help readers to follow. For example,

line 183, global, Atlantic and Pacific → (panel a) global, (b) Atlantic and (c) Pacific
line 290, Figure 2 → Figure 2(a)
line 370 Figure S9 → Figure S9 upper panel
line 371 Figure 10S → Figure 10S upper panel
line 372 Figure S9 → Figure S9 bottom panel
line 439, Figure 2 → Figure 2(b)
… and so on.

4. The sequence of figures in Supplement is suggested to follow the sequence of appearance in the main text. For example, Figure S12 appears in line 294, Figure S6 in line 314, Figure S4 and S5 in line 315, in the main text. It is suggested to re-order figures in Supplement.

5. Lines 441 and 447, Typo, SSP5-RCP8.5 → SSP5-8.5