Reviewer 1

I would like to thank the author for replying to my comments.
From the reply it seems like, maybe, I was not clear enough about my main concern. If that was the case, I am sorry.

The author demonstrates, using the standard error of 60 years slab ocean simulations, that (at least in some of the cases) simulations with added anthropogenic aerosols at a given location are different from a reference simulation, which include no anthropogenic aerosols. I have no concern about this part. My concern is about the attribution of the difference to the added aerosols. Two different 60 year-long simulations could by statistical different from each other (in a 95% confidence level) even without any external forcing, i.e., only due to internal variability. Based on the presented results in this paper, one can't rule out that this is the case here. This simply cannot be done in a relatively short (60 years) single simulation for each forcing pattern (unless the signal is orders of magnitude larger than the range possible due to internal variability, which is not the case here, even compared to slab ocean simulations and not fully coupled simulations).
In order to robustly attribute precipitation changes to aerosols, large ensemble of simulations or very long (1000's years) simulations are needed. Hence, I would still like to encourage the author to conduct, at the very least, one or two more simulations for each aerosol location (with slightly different initial conditions). This could at least strength the attribution argument, even though without a large ensemble (few 10's) of simulations it will not completely rule-out the role of internal variability.
Finally, a word about the comparison of slab ocean and fully-coupled ocean simulations. No doubt that the estimation of the statistical significance, as well as of the role of the forced response (compared to natural variability) should be derived from a similar dataset. However, the atmospheric natural variability is still very high in slab ocean simulations and the range of precipitation due to natural variability alone is not expected to be much smaller than in coupled simulation. In fact, even in simulations with prescribed SST a large-ensemble is needed many times to identify atmospheric response (see for example (Gervais et al., 2019)).

Gervais, M., Shaman, J., and Kushnir, Y.: Impacts of the North Atlantic warming hole in future climate projections: Mean atmospheric circulation and the North Atlantic jet, Journal of Climate, 32, 2673-2689, 2019.

I thank the reviewer for further clarifying their concerns regarding the methodology used in this study. Despite the fact that the methodology and significance testing applied in this study to identify climate responses attributable to changes in regional aerosol emissions is standard across the aerosol-climate literature (see below), I appreciate the opportunity to advance best practices and have made several improvements to the manuscript to alleviate any concerns regarding the robustness of the signals assessed.

- All slab ocean coupled simulations have been repeated with slightly adjusted initial conditions. All analysis conducted on the original simulation set is now repeated in this second simulation set, and comparable figures for the second simulation set are now included in Appendix A. The results from the second simulation set are consistent with the original simulation set, indicating that the signal is unlikely to arise from internal variability.
- The concern raised by the reviewer that “two different 60 year-long simulations could by statistical different from each other (in a 95% confidence level) even without any external forcing, i.e., only due to internal variability” could emerge if a persistent mode of internal variability was present in the perturbation simulation and thereby conflated with the perturbation signal. This is addressed by the additional simulations described

above. However, it was also addressed in the original submission by adjusting the effective sample size of all statistical tests for autocorrelation between years. Any persistent mode of internal variability would be expected to increase autocorrelation between years. Were a persistent mode to produce the signals evaluated, it would be expected to reduce the effective sample size to the point where significant signals would not be detected. Following best practices (e.g. Westervelt et al. 2020, Conley et al., 2018), I adjust the effective sample size to account for autocorrelation following the methodology of Santer et al. (2000). This practice is now described more clearly in the Methodology section of the manuscript.

- I would like to note that I use the 95% confidence interval to characterize uncertainty throughout, not the standard error as the reviewer initial stated.

Results of the new simulation set are provided in Figures A2-A5 and the above improvements to the manuscript are detailed in Section 2, L119-144 of the revised manuscript as follows:

"The simulation design used here, in which a signal from a given perturbation (e.g. a regional aerosol emissions) is characterized by imposing that perturbation as the only modification to a control simulation and running the resulting simulation in repeating annual cycle mode for an extended period, is a standard methodology used across the aerosol-climate interactions literature. Examples include simulations conducted as part of the Precipitation Driver and Response Model Intercomparison Project (PDRMIP) (e.g. L. Liu et al., 2018; Myhre et al., 2016; Samset et al., 2016) with idealized regional aerosol perturbations and within multi-model (Westervelt et al., 2017, 2018) and single-model experiment designs (Kasoar et al., 2018) simulating removal of present-day aerosol emissions in individual regions. In this experiment design, the perturbation signal is characterized as the difference between the long-term mean of the perturbation and control experiments after they have reached quasi-equilibrium, and the effects of internal variability are estimated using the interannual variability between individual years of the simulation.

One concern with this approach, not addressed in prior studies, is that persistent modes of internal variability may emerge within the equilibrium simulations and could be conflated with the perturbation signal. While atmosphere-only simulations cannot sustain long-term modes of internal variability, this concern may apply to the slab ocean coupled simulations used here. To address this concern, two approaches are applied in this study. First, statistical significance is estimated using either the last 60 years (slab ocean simulations) or 40 years (atmosphere-only) of the simulations as the sample, but effective sample size is adjusted to account for autocorrelation between simulation years following the methodology of (Santer et al., 2000). The 95% confidence level (i.e. 1.96σ) based on year-to-year variability in the difference between the control simulation and each perturbation experiment is provided for all global-mean values and statistical significance for maps is estimated at the 95% confidence level using a two-tailed *t* test, both using this adjusted effective sample size. Second, the slab ocean coupled experiments are repeated with slightly adjusted initial conditions (initial conditions drawn from a different year of the control simulation), allowing a different trajectory of internal variability to emerge within the equilibrium simulation. Results from this second experiment set are provided in Appendix A and demonstrate that the central findings of the study are unlikely to be the result of persistent modes of internal variability emerging in either equilibrium simulation set, but rather can robustly be assumed to result from the regional aerosol emissions perturbations imposed."

I would, however, like to highlight that the methodology and significance testing applied in this study to identify climate responses attributable to changes in regional aerosol emissions is standard across the aerosol-climate literature. All of the below studies are conducted as repeating annual cycle equilibrium simulations, as in the current study. Several studies use perturbations comparable in magnitude to that imposed in this study (Westervelt et al., 2020, Westervelt et al., 2018, Kasoar et al., 2016, Kasoar et al., 2018) and run for a similar (O(100 years)) duration. Notably, none of them run multiple ensemble members for a given equilibrium simulation, as the purpose of the equilibrium simulation is to sample internal variability across the duration of the simulation. Thus, the reviewers statement that the signals evaluated in this study cannot be attributed to the aerosol perturbation without a much larger perturbation, a large ensemble, or an O(1000 year) equilibrium simulation are not supported by the existing literature.

In addition, all of the below studies use comparable statistical tests to the one used here to attribute the signal seen in their simulations to the imposed aerosol perturbation. Indeed, several do not conduct the correction for autocorrelation conducted in this study. See for example:
Westervelt et al., 2020: statistical significance at the 95 % level according to a Student t test with the false discovery rate method from Wilks (2016) applied and an effective sample size adjusted for autocorrelation used [methodology comparable to that used in this study].
Westervelt et al. 2018: Hatching represents statistical significance at the 95 % level according to a Student's t test.
Kasoar et al., 2016: Stippling indicates that the change in that grid box exceeded 2 standard deviations [i.e. approximately the 95% confidence interval].
Kasoar et al., 2018: stippling indicates that the change at that grid-point exceeded 2 standard deviations [i.e. approximately the 95% confidence interval].
Liu et al., 2018: Stippled regions indicate where the multi-model mean change departs from zero by more than one standard deviation [i.e. approximately the 67% confidence interval – a weaker statistical threshold than the one applied here].
Samset et al., 2016: Hatched regions indicate where the multi-model mean is more than 1 standard deviation away from zero. [i.e. approximately the 67% confidence interval – a weaker statistical threshold than the one applied here]
Zhang et al., 2021: Hatching indicates where the changes are "significant" (90 % confidence).

References cited:

Kasoar, M., Shawki, D. & Voulgarakis, A. Similar spatial patterns of global climate response to aerosols from different regions. *npj Climate and Atmospheric Science* **1**, 12 (2018).

Kasoar, M. *et al.* Regional and global temperature response to anthropogenic SO_2 emissions from China in three climate models. *Atmospheric Chemistry and Physics* **16**, 9785–9804 (2016).

Westervelt, D. M. *et al.* Connecting regional aerosol emissions reductions to local and remote precipitation responses. *Atmospheric Chemistry and Physics* **18**, 12461–12475 (2018).

Westervelt, D. M. *et al.* Local and remote mean and extreme temperature response to regional aerosol emissions reductions. *Atmospheric Chemistry and Physics* **20**, 3009–3027 (2020).

Samset, B. H., G. Myhre, P. M. Forster, Ø. Hodnebrog, T. Andrews, G. Faluvegi, D. Fläschner, et al. 2016. "Fast and Slow Precipitation Responses to Individual Climate Forcers: A PDRMIP Multimodel Study." *Geophysical Research Letters* 43 (6): 2016GL068064. https://doi.org/10.1002/2016GL068064.

Santer, B. D., T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor. 2000. "Statistical Significance of Trends and Trend Differences in Layer-Average Atmospheric Temperature Time Series." *Journal of Geophysical Research: Atmospheres* 105 (D6): 7337–56. https://doi.org/10.1029/1999JD901105.

Zhang, S., Stier, P. & Watson-Parris, D. On the contribution of fast and slow responses to precipitation changes caused by aerosol perturbations. *Atmos. Chem. Phys.* **21**, 10179–10197 (2021).

Reviewer 2

This is my second time viewing this paper and all my precius comments are satisfactory addressed.

I thank the reviewer for their favorable evaluation of the manuscript and for their earlier feedback, which contributed to improvements in the manuscript.