In this paper the author uses a series of single model simulations with variations in the aerosol emission's geographical location to study the dependence of precipitation on aerosol emission location. The manuscript is well written, nicely organized and study interesting and important topic. However, I have a strong concern that many of the results are not significantly differ from changes one can get just due to internal variability. In the global mean, the changes presented here are in the range of ~ 0.002-0.02 mm/day. I have calculated the pre-industrial (PI) CESM1 global mean precipitation to be 3.044 mm/day, i.e., the changes seen here are in the range of 0.07-0.7% of the global mean. A question which is central to this paper is whether the changes reported here are statistically significant compared to the natural verbality of the system. The range of the 40-year running mean global mean precipitation in the CESM1 PI is 3.037 – 3.051 mm/day, i.e., a range of 0.014 mm/day just due to natural viability. That means that at least some of the difference in the global mean precipitation seen here are within the range possible only due to natural variability (all thought some of them are outside this range). Looking on the significant test in Fig. 1, suggests that the same might be true for the local precipitation changes (the changes in the vast majority of places aren't significant). In addition, statistically significant local precipitation variations between two realizations could also be driven by natural variability and not by the external (aerosol) forcing.

The way to overcome this issue, and to make sure that the differences are driven by the aerosol forcing and not by natural variability is to simulate more realizations (i.e., conduct initial-condition large-ensemble) (Diao et al., 2021), or run the model for very long times (Fiedler & Putrasahan, 2021). I feel bad to ask such a revision as I understand the amount of work it might require. However, I hope that, if the author will accept my suggestion, the paper could become much more convincing. If conducting a large-ensemble is beyond the reach of the author, I believe that conducting one or two more simulations for each aerosol location (with slightly different initial conditions) will improve the confidence in the results (in case they are similar to the initial results) or demonstrate the need in more realizations (in case they are not similar, thus suggesting a large role of natural variability).

I thank the referee for their thoughtful consideration of the results and for their favorable review of the overall manuscript structure and content. I agree that characterizing the statistically significant precipitation response to aerosol perturbations is a difficult process. However, I believe the referee's concerns regarding statistical significance are based on a misapprehension. I detail this below, as well as steps that are taken in the revised manuscript to make the reliability of the signals clearer. I nevertheless appreciate the interrogation of the statistical significance calculation, as this is an important factor to clarify.

The PI CESM1 variability the referee estimates appears to be from the fully coupled CESM1, which will have different variability than the slab-ocean set-up used in this study (see Section 2). Measures of statistical significance should be derived from the native dataset. I use 60 years of simulations in the slab-ocean configuration to characterize the precipitation signal. Standard errors derived from the interannual variability in the difference between the slab ocean PI control and each perturbation experiment were previously given in Figure 1. The standard error values may be roughly doubled to provide an estimate of the 95% confidence range. This 95% confidence range is also provided as error bars on the fast and slow precipitation responses in Figures 2 and 3 (note that this information has now been added to the relevant figure caption). <u>From this significance calculation, it is clear that for all regional perturbations aside from Indian and East African emissions the global-mean total precipitation response is statistically distinguishable from zero at the 95% confidence level after accounting for internal variability and thus is highly unlikely to have arisen from internal variability alone.</u> (The global-mean precipitation response to East African and Indian emissions are statistically indistinguishable from each other and from zero.)

<u>Furthermore, the paper does not claim that every global-mean response is statistically distinct from the others, nor does its argumentation depend on this.</u> There is clearly statistically significant diversity in the global-mean response to identical aerosol emissions from different regions (a range of -2.6 to -21.3 μm/day compared to a maximum standard error of 1.8 μm/day and a maximum 95% confidence interval of 3.6 μm/day), which forms one of the motivations for the analysis in the paper.

Regarding the statistical significance shown on the maps, it is to be expected that regional aerosol perturbations will produce spatially heterogeneous precipitation changes that will be statistically significant in some regions and not in others. The spatial extent of regions with statistical significance is comparable or higher than that seen in other modeling studies of regional aerosol effects on precipitation using forcings of similar magnitude across several different climate models and a similar statistical significance criterion. For examples, see the following:

- Westervelt et al., 2017, Figure 3 – fully coupled 200-year simulations of removal of U.S. sulfate emissions in NCAR CESM, GFDL CM3, and GISS E2 models
- Westervelt et al., 2018, Figure 1 – fully coupled 200-year simulations of removal of U.S., European, and Asian sulfate, bc, or all anthropogenic aerosol emissions in NCAR CESM, GFDL CM3, and GISS E2 models
- Kasoar et al., 20218, Figure S2 – fully coupled 150-year simulations of removal of North American, European, East Asian, or South Asian sulfate emissions in HadGEM3-GA4

Westervelt, D. M., Conley, A. J., Fiore, A. M., Lamarque, J.-F., Shindell, D., Previdi, M., Faluvegi, G., Correa, G., & Horowitz, L. W. (2017). Multimodel precipitation responses to removal of U.S. sulfur dioxide emissions. *Journal of Geophysical Research: Atmospheres*, *122*(9), 2017JD026756. https://doi.org/10.1002/2017JD026756

Westervelt, D. M., Conley, A. J., Fiore, A. M., Lamarque, J.-F., Shindell, D. T., Previdi, M., Mascioli, N. R., Faluvegi, G., Correa, G., & Horowitz, L. W. (2018). Connecting regional aerosol emissions reductions to local and remote precipitation responses. *Atmospheric Chemistry and Physics*, *18*(16), 12461–12475. https://doi.org/10.5194/acp-18-12461-2018

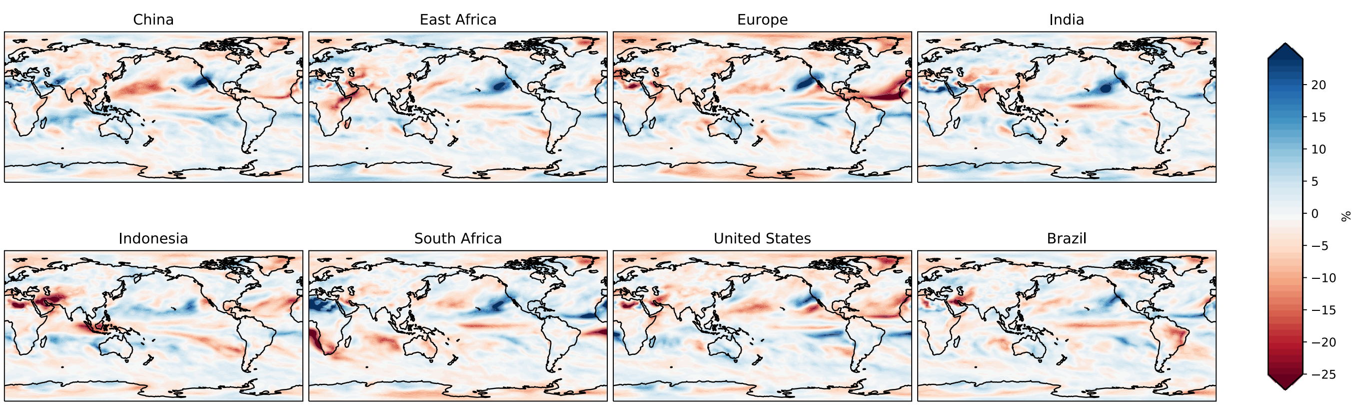
Kasoar, M., Shawki, D., & Voulgarakis, A. (2018). Similar spatial patterns of global climate response to aerosols from different regions. *Npj Climate and Atmospheric Science*, *1*(1), 12. https://doi.org/10.1038/s41612-018-0022-z

The following improvements will be made to the revised manuscript:

- I now report the 95% confidence interval rather than the standard error on Figure 1 and wherever the global-mean total precipitation responses are reported in the text. Error ranges given on Figure 5 (fast precipitation response) and Figure 3 are now also the 95% confidence interval rather than the standard error as well. All error ranges throughout the manuscript thus now provide the 95% confidence interval, allowing clear depiction of whether all signals are is distinguishable from natural variability with high confidence.
- I summarize the above discussion of statistical distinctness of the global-mean precipitation response as follows (L139-145):
  "The global-mean precipitation response to Indian and East African emissions, which constitute the weakest of the precipitation responses, are statistically indistinguishable from zero and from each other in the presence of internal variability. All other global-mean precipitation responses are statistically significant at the 95% confidence level, and thus highly unlikely to arise from internal variability alone. Although the 95% confidence interval in the global-mean response to some regional emissions are overlapping, it is clear that there is statistically significant diversity in the global-mean response to identical aerosol emissions from different regions."
- I now explicitly indicate on Fig. 4a whether the regional-mean precipitation responses to in-situ aerosol changes are statistically significant at the 95% confidence level.
- I now improve the density and visibility of the statistical significance masking on all map figures in the main text (Figures 1, 5, and 6).

In addition, I believe that presenting Fig. 4 in relative terms will be more appropriate as the difference in the background precipitation between these places is very large.

I now include the precipitation response as a percent of the climatological precipitation in each grid box in the supplementary materials (I also include it below). As already stated in the manuscript, the distinction between regional emissions producing strong versus weak in-situ precipitation responses holds true whether the precipitation responses are considered in terms of absolute values or percent values. However, I have chosen to keep the absolute precipitation change as the format for the precipitation maps in the main manuscript, as this allows a more accurate portrayal of the spatial pattern of precipitation change (percent values may amplify the appearance of the precipitation response in some regions, if the denominator is small). It is for this reason that absolute precipitation changes are the standard mode of depiction of spatial patterns of precipitation response across recent papers exploring precipitation responses to regional aerosols, e.g. Westervelt et al. (2017), Westervelt et al. (2018), and Kasoar et al. (2018) cited previously, as well as Liu et al. (2018) and Samset et al. (2016).



Liu, L., Shawki, D., Voulgarakis, A., Kasoar, M., Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Ø., Sillmann, J., Aalbergsjø, S. G., Boucher, O., Faluvegi, G., Iversen, T., Kirkevåg, A., Lamarque, J.-F., Olivié, D., Richardson, T., Shindell, D., & Takemura, T. (2018). A PDRMIP Multimodel Study on the Impacts of Regional Aerosol Forcings on Global and Regional Precipitation. *Journal of Climate*, *31*(11), 4429–4447. https://doi.org/10.1175/JCLI-D-17-0439.1

Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Ø., Andrews, T., Faluvegi, G., Fläschner, D., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J.-F., Olivié, D., Richardson, T., Shindell, D., Shine, K. P., Takemura, T., & Voulgarakis, A. (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