Referee #2:

In " Contrasting Impacts of Asian Sulfate Aerosols on Regional Summer Extreme Precipitation under Different Climate Backgrounds ", Feifei Luo et al. use the coupled climate model FORTE2 to investigate the effects of Asian sulfate aerosols on regional summer extreme precipitation under two climate backgrounds: a pre-industrial climate state and a climate state 1 K warmer than pre-industrial conditions. The authors also compare FORTE2 results with simulations from the more comprehensive PDRMIP simulations. Overall, this is a valuable study that addresses an important and timely topic and has the potential to be of broad interest. However, the manuscript requires revisions before it can be considered for publication. In particular, it would benefit from substantial improvements in clarity, structure, and presentation. Many sentences are overly long and difficult to follow, which at times obscures the main message.

We sincerely thank you for your constructive and insightful feedback, which has greatly helped us improve the quality of our manuscript. We have carefully considered all the comments and revised the manuscript accordingly. We hope that the revised version and our point-by-point responses adequately address all the concerns raised. The main revisions and our detailed responses are presented below.

In addition, the model evaluation lacks a direct and consistent comparison framework, as the simulations being compared are conducted under different forcing and background climate conditions. This makes it difficult to assess the robustness of the FORTE2 results and weakens confidence in the conclusions. The authors should better justify the comparability of the evaluated simulations and provide a clearer assessment of model performance and result robustness.

Response: Thank you for raising this important issue. We fully agree that a direct and consistent comparison with historical simulations forced by observed external forcing would, in principle, provide a more useful benchmark and strengthen confidence in model evaluation. However, such a simulation is not feasible with FORTE2, as the ACI is tuned and added separately. In addition, due to the relatively low transient climate response of FORTE2, a simulation using present-day well-mixed greenhouse gas concentrations would produce an SST warming that is substantially smaller than the observed present-day conditions. Hence, the p1K simulation (CO_2, 500ppmv) was designed to provide an SST state more representative of present-day conditions and serves as the baseline for the aerosol perturbation simulations.

Therefore, this evaluation is not intended as a formal validation of the model's ability to reproduce the absolute observed climate. Rather, it serves as a sanity check to ensure that the model generates reasonable precipitation and physically plausible large-scale circulation under its own boundary conditions, which is a necessary prerequisite for conducting any meaningful perturbation experiment. We have clarified this point at the beginning of the section 3.1 (Model evaluation) in the revised manuscript, and revised this section accordingly to emphasize the comparison of spatial structures rather than the exact magnitudes of the climatological fields. Please see the revised text in lines 267-311.

We further assess the robustness of the simulated aerosol–precipitation responses by comparing FORTE2 and PDRMIP in Section 3.2 (Responses to Asian sulfate aerosols and associated physical processes). The response is defined as the difference between the perturbation and corresponding baseline simulations. Importantly, the aerosol forcing used in both FORTE2 and PDRMIP are idealized and designed to isolate the effects of sulfate aerosol forcing. As shown in Fig. 1, the spatial patterns of the Asian sulfate AOD forcing are similar between the two experiments, although the PDRMIP forcing is approximately 10 times stronger than that in FORTE2. The stronger forcing in PDRMIP is intended to amplify the precipitation response and facilitate a clearer identification of the underlying physical mechanisms. Therefore, the two experiments are not expected to produce identical response magnitudes, and our comparison focuses instead on the sign and spatial characteristics of the responses. We have clarified this point in the revised manuscript. Please see lines 205-211. The two models show consistent response signs and spatial patterns, as well as similar underlying physical mechanisms, in response to Asian sulfate aerosol forcing. The simulated precipitation responses are also consistent with previous sulfate aerosol perturbation experiments using other models (Liu et al., 2009; Guo et al., 2013; Westervelt et al., 2020). Together, these consistencies across models and previous studies provide further support for the robustness of the simulated responses and associated mechanisms.

Liu, Y., Sun, J., and Yang, B.: The effects of black carbon and sulphate aerosols in China regions on East Asia monsoons, Tellus B, 61(4), 642-656, 2009

Guo, L., Highwood, E. J., Shaffrey, L. C., and Turner, A. G.: The effect of regional changes in anthropogenic aerosols on rainfall of the East Asian Summer Monsoon, Atmos. Chem. Phys., 13, 1521–1534, https://doi.org/10.5194/acp-13-1521-2013, 2013.

Westervelt, D. M., You, Y., Li, X., Ting, M., Lee, D. E., and Ming, Y.: Relative importance of greenhouse gases, sulfate, organic carbon, and black carbon aerosol for South Asian monsoon rainfall changes, Geophys. Res. Lett., 47(13), e2020GL088363, 2020

Although a Results section and a separate Discussion section is provided later, several subsections move prematurely into interpretation and discussion without adequately presenting the underlying results. I recommend first describing the key findings shown in the figures and then reserving interpretation and broader implications for the Discussion section. This would improve the logical flow and make it easier for readers to distinguish between the reported results and their interpretation.

The manuscript structure would also benefit from placing the Discussion before the Conclusion. Furthermore, the current Discussion section reads more like a recommendation than a discussion of the study’s findings. Consider expanding the discussion of the key results, their implications, limitations, and relation to previous studies, while reserving recommendations and future work for the concluding remarks.

Response: Thank you for this helpful suggestion. Accordingly, we have revised the Results section to present the key results shown in the figures before providing their broader interpretation. Further details are provided in our response to the related comment below.

Following your recommendation, we have placed the Discussion section before the Conclusion. We have also restructured and substantially revised the Discussion section to focus on the scientific interpretation and implications of the main findings, rather than recommendations. The revised Discussion now addresses (1) the dependence of the sulfate aerosol–precipitation response on the background climate state, (2) the underlying thermodynamic and dynamical mechanisms, (3) the climate-state dependence of ARI–ACI interactions, (4) consistency with PDRMIP results and previous studies, (5) the limitations of the FORTE2 framework, and (6) the potential implications for ongoing and future sulfate aerosol reductions. Please see lines 473-536.

The Conclusion has also been revised to provide a concise synthesis of the main findings and their key implications, while recommendations and future research directions are kept brief and presented at the end. Please see lines 538-571.

Line 42: The statement is somewhat ambiguous. Please clarify that the aerosol-induced reduction in precipitation is enhanced under a warmer climate background, rather than implying that global warming itself strengthens the aerosol effect.

Response: Thank you for pointing this out. We have rephrased this sentence. Please see lines 41-43: "This study suggests that the reduction in precipitation, including extremes, over northern India caused by Asian sulfate aerosols is amplified under a warmer climate background."

Line 110: Could the authors provide more details on how the vertical layer thickness varies with depth? Additionally, please clarify the rationale and advantages of using progressively thicker layers at greater depths.

Response: Thank you for this question. The FORTE2.0 configuration used in this study employs 15 z-levels in the MOMA ocean model, with layer thickness increasing progressively with depth, from approximately 30 m near the surface to 800 m in the deepest layer (Blaker et al., 2021; Stjern et al., 2024). The total model ocean depth is approximately 5,000 m. However, neither Blaker et al. (2021) nor Stjern et al. (2024) provides further details on the thickness of individual layers, and the detailed layer interfaces are also not explicitly tabulated in the standard MOMA documentation (Webb, 1996). Therefore, we are unable to provide the exact thickness of each individual layer beyond the information available above. Such a progressively increasing layer thickness provides finer vertical resolution in the upper ocean, where the mixed layer and air–sea interactions exhibit stronger vertical gradients, while coarser resolution at greater depths is appropriate because oceanic properties generally vary more gradually with depth and helps reduce computational cost.

Lines 138–139: Could the authors explain the rationale for imposing the constraints σmin < 0.75 and pmin > 300 hPa over complex terrain? In particular, it would be helpful to describe their impact on the model performance and to provide references supporting the choice of these threshold values.

Response: Thank you for this comment. These constraints are imposed to avoid unrealistic aerosol vertical distributions over complex terrain, following the aerosol forcing configuration described by Stjern et al. (2024). σmin< 0.75 prevents the aerosol layer from extending too close to the surface. The pmin>300hPa prevents thin aerosol layers with spuriously high AOD from being placed in the upper troposphere over mountainous terrain. In practice, these thresholds are rarely triggered: most pmin values are around 600 hPa over South Asia and 700 hPa elsewhere (Fig. 1), suggesting that their influence on the prescribed aerosol distribution is limited.

We have added the reference to Stjern et al. (2024) and clarified the rational of these constraints in the revised manuscript (Lines 144–146): “In areas with complex terrain, additional constraints are enforced to avoid unrealistic aerosol vertical distributions, with σ_{min} constrained to below 0.75 and p_{min} to above 300 hPa (Stjern et al., 2024).”

Stjern, C., Joshi, M., Wilcox, L., Gollo, A. and Samset, B.: Systematic Regional Aerosol Perturbations (SyRAP) in Asia using the intermediate-resolution global climate model FORTE2, J. Adv. Mod. Earth System, 16, e2023MS004171, 2024.

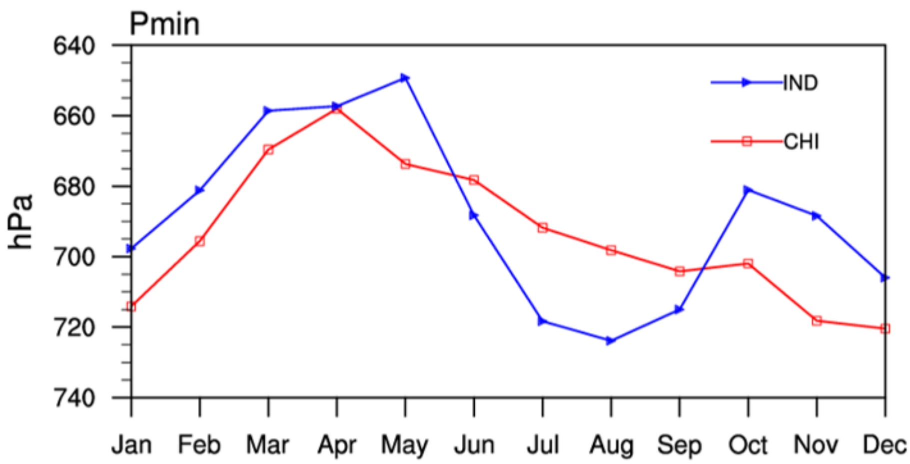

Figure 1. Seasonal evolutions of the regional mean p_{min} for China and India.

Lines 149–155: are somewhat lengthy and difficult to follow. Please consider rephrasing this section to improve clarity and readability.

Response: Thank you for this helpful suggestion. We have revised the paragraph to improve clarity and readability by splitting it into shorter sentences and clearly separating the baseline and perturbation simulations. Please see lines 156-163: “The

FORTE2 simulations used in this study are summarized in Table 1. They include two baseline simulations without aerosol forcing: one with CO_2 concentrations at pre-industrial levels (piC, 280 ppmv), and one with elevated CO_2 (p1K, 500 ppmv). In this model, the p1K simulation yields a global-mean warming of approximately 1K relative to piC. In addition, six perturbation simulations were performed with SO_4 AOD forcing added over Asia (Fig. 1a). These comprise combined ARI and ACI effects (ARI+ACI), ARI-only and ACI-only effects, each run under both the piC and p1K background climate states."

Line 155 – 156: The wording is unclear. Do the authors mean that the first 50 years were treated as a spin-up period and excluded from the analysis? If so, please state this explicitly.

Response: Thank you for pointing out this ambiguity. We have rephrased the relevant statements. Please see lines 163-165: "All simulations were carried out over a total 200-year period, with the first 50 years discarded as spin-up, and the remaining 150 years (model years 51-200) used for the analysis."

Fig. 1: The two panels are presented using different scales, which makes a direct comparison difficult. The authors state that the figure is intended to highlight differences in spatial patterns and signs; however, most regions appear positive in both panels. Would the authors consider including a direct comparison, such as a normalized difference or relative change, to better illustrate the differences?

Response: Thank you for this insightful comment. We fully agree that the use of different color scales in the two panels of Fig. 1 may hinder direct comparison. However, we would like to clarify that Fig. 1 was designed to serve two complementary purposes. First, as stated in the manuscript, we aim to emphasize the magnitude of the prescribed SO_4 perturbation is much stronger in PDRMIP than in FORTE2, by roughly an order of magnitude. Hence, if a single unified color scale were used, the positive AOD signals in the FORTE2 panel would be nearly invisible against the background, making the spatial pattern in FORTE2 barely discernible.

Second, beyond the magnitude contrast, we also intended to compare the spatial patterns between the two datasets. Despite the large difference in absolute values, both simulations show similar spatial distributions, with high AOD centers located over East Asia and South Asia, especially over North China. We acknowledge that the original text focused only on the magnitude difference and did not explicitly state the intention of comparing spatial patterns. In the revised manuscript, we have clarified this point, please see lines 205-211: "It should be noted that the PDRMIP simulations prescribe a much stronger Asian SO_4 perturbation than FORTE2, with a regional summer multi-model mean (MMM) AOD of ~2.5 over East Asia and South Asia, compared to ~0.23 in FORTE2 (Fig. 1). Despite this difference in magnitude, the spatial pattern of the prescribed perturbation is similar between the two experiments (Fig. 1). Therefore, the comparison between PDRMIP and FORTE2 focuses on the sign and spatial pattern of the responses, rather than their magnitudes, in the subsequent analysis."

We also considered the reviewer's suggestion of presenting a normalized difference or relative change. However, such a representation would obscure the large difference in absolute magnitude between the two simulations, which is one of the key points of Fig. 1. Therefore, we have retained the current format with separate color scales, as it allows readers to clearly see both the spatial distribution and the magnitude contrast in each dataset.

Lines 193–197: The interpretation is largely left to the reader. Rather than only stating that the sign and spatial pattern can be compared, the authors should provide their own assessment of the model performance and discuss what the comparison with PDRMIP implies about the robustness and limitations of the FORTE2 results.

Response: Thank you for pointing out this issue. We agree that the original text in Section 2.2 included a statement about the robustness of the FORTE2 results, which was indeed interpretive and not appropriate for a methods section. In the revised manuscript, we have removed this statement "which would support the robustness of the FORTE2 results", and replaced it with an objective description of the comparability between PDRMIP and FORTE2, namely that the comparison should focus on the sign and spatial pattern of the responses rather than their magnitude, owing to the much stronger prescribed SO_4 perturbation in PDRMIP. In addition, a detailed assessment of model performance is provided in Section 3.1 (Model evaluation) and Section 3.2 (Responses to Asian sulfate aerosols and associated physical processes), and the limitations are further discussed in Section 4.

Section 2.1: **SyRAP-FORTE2** is discussed as a version of the FORTE2 model; however, the term **SyRAP-FORTE2** rarely appears beyond the model description and is not used consistently in the Results, Discussion, or Conclusion sections. It is therefore unclear whether the reported results are produced by **SyRAP-FORTE2**, **FORTE2**, or whether the two names are being used interchangeably. The authors should clarify the relationship between these model versions and consistently indicate which model is used throughout the manuscript to avoid confusion.

Response: Thank you for raising this important clarity issue. In the revised manuscript, we have removed all instances of "SyRAP-FORTE2" and now use "FORTE2" consistently to refer to the simulations conducted in this study. The revised text in Section 2.1 now reads as follows (lines 119-121): "Building on the FORTE2 platform, Stjern et al. (2024) designed and implemented a suite of Systematic Regional Aerosol Perturbations experiments. For brevity, hereafter we refer to these simulations collectively as FORTE2."

Lines 160–167: The large number of correlation coefficients is difficult to follow. Could the authors provide a brief summary of the key results, as a table?

Response: Thank you for this helpful comment. We suspect the reviewer may have been referring to lines 260–267, rather than the lines 160-167 in the comment, where several correlation coefficients were originally presented. The revised text as follows (lines 286-292): “FORTE2 also reasonably simulates these precipitation indices, although its pattern correlations are generally lower than those of the PDRMIP MMM, with values of 0.56 for PRCPTOT and 0.63 for R95p (Fig. 2i-j). Notable dry biases are also evident over India for PRCPTOT and R95p, a feature common among many CMIP5/6 models (Wilcox et al., 2020). The correlation is lower for R95d (0.39), largely associated with its relatively low values in the tropics, while R95int shows a comparatively higher correlation of 0.68 (Fig. 2k-l).”

Since the specific correlation values are already labeled in Fig. 2, we believe an additional table is not necessary. However, if you still prefer a table, we would be glad to provide one.

Lines 260–262: Could the authors provide the corresponding figure or table of reference for these values to help readers locate the supporting results?

Response: We appreciate the suggestion. Rather than adding a separate figure or table, we have clarified this point in the text as a general conclusion, given that this individual model correlations serve only as brief background context rather than a central result of this study. Moreover, since the subsequent analyses are based primarily on the MMM, the detailed values for individual models are not essential to the main conclusions.

The revised text now reads (lines 284-286): “All individual PDRMIP models show pattern correlations above 0.6 for these indices, indicating reasonable agreement with the observed spatial patterns.”

Line 258 and 263: "PDRMIP" and "MMM" appear to be used interchangeably. Please use consistent terminology to avoid confusion.

Response: Thank you for this helpful comment. We have carefully checked the manuscript and revised the terminology throughout to consistently use “PDRMIP MMM” when referring to the multi-model mean of the PDRMIP models.

Lines 272 – 274: The model evaluation appears to compare simulations conducted under substantially different forcing setups and background climate states. For example, FORTE2 includes aerosol-free and P1k experiments, whereas the PDRMIP baseline is configured with present-day aerosol and greenhouse gas forcing. Could the authors justify why these simulations are considered sufficiently comparable for model evaluation?

Response: Thank you for raising this point again. We agree that the FORTE2 and PDRMIP simulations are not conducted under identical forcing and background climate conditions, and therefore should not be interpreted as a direct validation against observations. The FORTE2 p1K simulation was designed to provide an SST state more representative of present-day conditions, given the relatively low transient climate response of FORTE2, while the PDRMIP baseline represents year-2000 aerosol and greenhouse gas forcing. In addition, the aerosol distributions prescribed in both FORTE2 and PDRMIP are idealized sulfate perturbations designed to isolate the effects of Asian sulfate aerosols, rather than to reproduce the full historical aerosol distribution.

Accordingly, we use the comparison primarily as a sanity check of whether the models capture the major spatial structures of precipitation and large-scale circulation relevant to the Asian summer monsoon, rather than as a comparison of the absolute magnitudes of their climatological fields. The differences in forcing and background state can affect the magnitude and regional details of the simulated climatology, but do not preclude a comparison of their major spatial characteristics. We have clarified this rationale at the beginning of Section 3.1 (lines 267-277).

The comparability of the aerosol-induced responses is further assessed in Section 3.2. Although the PDRMIP sulfate forcing is approximately ten times stronger than the FORTE2 forcing, the spatial patterns of the prescribed Asian sulfate AOD are similar between the two experiments. We therefore focus on the sign and spatial patterns of the precipitation responses rather than their magnitudes. The consistency of these responses and their dominant physical mechanisms between FORTE2 and the PDRMIP models, together with agreement with previous sulfate aerosol perturbation studies, provides additional support for the robustness of the results.

Line 297: Please clarify what is meant by "shown here". Consider providing an explicit cross-reference to the relevant figure, table, or section.

Response: Thanks for this suggestion. We have revised the text to replaces the vague phrase. The sentence now reads (lines 315-317): “To intuitively reflect the magnitude of regional precipitation responses, relative changes in PRCPTOT, R95p and R95int with respect to the respective baseline simulations are shown in Figs. 4, 7-8, 11 and 13.”

Line 316: "East Asia and South Asia**"** A cross-reference to "Fig. 4a" would be helpful here, allowing readers to easily identify the regions discussed and relate the text to the spatial patterns shown in the figure.

Response: Following your suggestion, we have added a cross-reference to Fig. 4a to help readers identify the regions discussed. The sentence now reads (line 337): “… East Asia and South Asia (red and purple boxes in Fig. 4a, respectively).”

Section 3.3: This section appears to move directly into interpretation and discussion without first presenting the results shown in the individual figures. I suggest first describing the main findings from Figs. 7 - 10 separately and then discussing their implications. As currently written, readers are expected to infer the results.

Response: Thank you for this helpful suggestion. In the revised manuscript, we have restructured this section so that each figure is now introduced with a dedicated sentence summarizing its main findings before any discussion. Please see the section 3.3.1 in the revised manuscript (lines 377-425).

Section 3.4: Similar to Section 3.3, this section moves rapidly into interpretation without adequately presenting the results shown in Figs. 11–13. Please consider first summarizing the main findings from each figure before discussing their implications.

Response: Thank you for this comment. In the revised manuscript, we have restructured this section so that each figure is introduced with a concise summary of its main findings before any discussion. Please see the section 3.3.2 in the revised manuscript (lines 427-472).

Lines 360–361: Since this is already the **Conclusion** section, the phrase "Our main findings are summarized as follows" is unnecessary and could be omitted for conciseness.

Response: We have removed the phrase "Our main findings are summarized as follows" from the Conclusion section as recommended.

Lines 503–504: The statement " This study underscores the strong dependence of aerosol-precipitation relationships on the background climate" appears somewhat broader than what is directly demonstrated in the manuscript. Consider limiting this conclusion to the specific regions and climate scenarios examined in this study, rather than implying a more general relationship.

Response: Thank you for raising this point. In the revised manuscript, we have narrowed the conclusion to the specific region and climate scenarios examined in this study. The sentence now reads (lines 474-475): "This study underscores the strong dependence of sulfate aerosol–precipitation relationships over northern India on the background climate state."

Line 503 – 508: While it is important to acknowledge model limitations. The wording in this paragraph sounds significantly to undermine confidence in the model and the study's conclusions.

Response: Thank you for this constructive comment. We have revised the paragraph to better balance the discussion of model limitations with the confidence of the conclusions. The revised text now reads (lines 517-525): "Several limitations should be considered. FORTE2 does not include interactive wet deposition feedbacks, which may somewhat amplify the precipitation response to aerosols. In addition, the model exhibits a dry bias in climatological summer precipitation over India. Nevertheless, the consistency of the dominant response and dynamical mechanism with the PDRMIP models and previous sulfate aerosol perturbation studies supports the robustness of the major findings. The FORTE2 experiments also provide a useful framework for isolating the influence of background climate state, which cannot be directly assessed from the single-background-state PDRMIP experiments."

Line 113: "adjustments" to adjustment

Response: It has been revised. Please see line 118.

Line 153: “Fig. 1a” Is it the correct reference? Please verify the cross-reference.

Response: Yes, "Fig. 1a" is the correct reference.

Line 283: Please define the acronym "SST" at first use (i.e., "Sea Surface Temperature").

Response: The sentence containing “SST” has been removed from the revised manuscript. Therefore, the acronym “SST” is no longer used in this section.