Response to Referee #1:

We are very grateful to the referee again in providing helpful comments and guidance that have improved the manuscript. In this document, we describe how we have addressed the reviewer's comments. Referee comments are shown in black and author responses are shown in blue text.

Overview:

The authors have mostly addressed the concerns I raised in the previous review, with a few exceptions. My major concerns are outlined below, followed by line comments.

Response: Thank you very much for your positive evaluations. All the questions and concerns have been carefully answered and the paper has been revised accordingly.

Major concerns:

1) Consistency of communication of findings:

In my initial review, I had suggested the authors stipple (add dots to) regions where the models disagree, but the authors only made this change to one figure. The other figures have stippling where the models agree- these dots cover the colors/shading in the locations where the changes are agreed upon by the models, effectively visually masking the relevant information. I suggest consistently using stippling throughout (in other words: stipple on all maps where models disagree, not where they agree on some maps and where they disagree on other maps- this can be confusing to the reader if the stippling changes from figure to figure).

Response: As suggested, we have stippled on all maps where models disagree in the revised manuscript.

The authors added country borders/boundaries to some of the maps, but not all of the maps- why did the authors only include this information on some of the maps and not others? I find the political boundaries/borders helpful for identifying which regions are impacted in which country when the

authors discuss the findings in the text.

Response: As suggested, we added country borders/boundaries to all the maps in the revised manuscript.

The authors are still using a very low significance criteria for stippling on the maps- most of the models must 'agree on sign of change' as the significance test. As I stated in my last review, I think in addition to agreeing on sign of change, some sort of signal: noise (or coefficient of variation or similar) would be a more robust/effective metric, particularly in the maps showing warming in ssp370 vs ssp370lowntcf- most of the globe warms almost everywhere over land in both experiments, so showing where the models all agree it warms as an agreement metric isn't particularly useful in my opinion. Agreeing on magnitude of changes would be more robust and meaningful.

Response: Thanks for your suggestion. Agreeing on sign of change was widely used in multi-model significance testing (Samset et al., 2018; Sillmann et al., 2019; Li et al., 2021). Also, as suggested, we added the information of signal-to-noise ratio, defined as the ratio of the absolute value of the multi-model ensemble (MME) mean to the deviation among the models, in the Supplements and corresponding description in the text to further describe the confidence level of the multi-model results. In addition, we also implemented statistical testing on the MME result (figures not shown). The regions with significance at ≥95 % confidence level from the t-test were broadly in line with those with signal-to-noise ratio larger than 1. So we only reserved the results of the signal-to-noise ratio in the revised manuscript.

2) Significance and uncertainty of findings, especially as they relate to precipitation changes.

In the text, I had suggested in my previous review that the authors present not only the multi-model mean/median when reporting findings, but also the spread or +/- 1 or 2 sigma. I appreciate the authors added this information in the Abstract and in some locations in the text, but they left out the uncertainty/model spread in many locations in the text. I suggest showing this uncertainty/spread throughout, both for consistency and transparency in terms of model agreement.

Response: As suggested, we added the corresponding uncertainty/spread in the text in the revised manuscript.

The authors have chosen to report several findings related to precipitation changes in the text with minimal discussion/mention (mentioned maybe once in the Discussion?) of the fact that the maps indicate most of the models don't even agree on the sign of the change (let alone the magnitude). Several of the precipitation results are reported in the text that seem to not be stippled in the maps, suggesting to me that the models don't agree, and I question how much we should trust the results. If the models don't agree on the sign of change in a region, I strongly suggest the authors mention this along with the results to warn the reader that the results should be interpreted with caution.

Response: As suggested, we mentioned that the sign of results shows poor agreement among models for the projection of extreme precipitation, especially in sub-regions in the revised manuscript as follows:

"Under the SSP3-7.0 scenario, heavy precipitation days decreases by approximately 2 days in central India, the Indo-China Peninsula, and by more than 4 days in the southeastern Qinghai-Tibet Plateau, SWC, and parts of Indonesia, while it increases in all other regions of the entire study area (Fig. 10a). However, it is worth noting that the consistency among models is poor." Lines 307-310

"In addition, increases of more than 3.0 days in heavy precipitation days are also found in India, the Indo-China Peninsula, and Indonesia, but again, consistency in these regions among models is poor, and this pattern is exactly opposite of that of GHG forcing (Fig. 10a and c)." Lines 314-316.

"In addition, non-methane SLCF reductions cause an increase of more than 2 mm in the maximum consecutive 5-day precipitation in the southeastern Qinghai-Tibet Plateau, the western Hengduan Mountains, SWC and the middle and lower reaches of the Yangtze River, whereas decreases of more than 3 mm occur in the southwestern part of the Indo-China Peninsula. However, there are large differences among models (Fig. 10f, Fig. S5f)." Lines 325-328

"Additional increases in total wet-day precipitation caused by reduced emissions of non-methane SLCFs are apparent in the southeastern Qinghai-Tibet Plateau, SWC, and the middle and lower reaches of the Yangtze River (>40 mm), while decreases occur in the southwestern part of the Indo-China Peninsula (<50 mm) with lower consistency among models." Lines 336-339

"However, consecutive dry days increases in NIN, the southeastern Tibetan Plateau, and the southern Yangtze River (Fig. 10j and k) with low inter-model agreement. Overall, consecutive dry days decreases by 0.02 ± 0.8 days and 0.4 ± 0.5 days under the SSP3-7.0 and SSP3-7.0-lowNTCF scenarios, respectively, with SNR less than 1, which means model projections have a low level of confidence. Future reductions in non-methane SLCFs contribute to decreases in consecutive dry days in northwestern and northeastern China, near the Hengduan Mountains, and the Indo-China Peninsula. Consecutive dry days decreases by 1.4 ± 1.2 days in SWC due to increase in the frequency and intensity of heavy precipitation (Fig. 11e). The sign of the results shows relatively good agreement and the SNR is larger than 1 among models in SWC and NC (Fig. 10i and Fig. S5i). Similarly, the changes in consecutive dry days due to non-methane SLCF reductions in some regions are comparable to the impacts of GHG forcing. Although for most regions, the models do not all agree on the sign of the responses, however, some robust features are still apparent." Lines 344-353

Related to this, the authors use wording such as 'significant increases are found' - If the authors are going to suggest that the results are 'significant', what is the significance test? If no test is used, can the authors use a test, such as a Mann-Whitney/Rank Sum test to determine if the medians/distributions are statistically distinguishable?

Response: Thanks for your constructive suggestion! We removed or replaced the word 'significant' and modified these sentences in the revised manuscript.

3) Proofreading/changing text after authors made changes:

There are several instances where the authors made changes to figures, but the figure numbers did not change (for example, there are two Figure 2s). Also, colors in line plots are changed from green to grey, but all captions still reference green lines that no longer exist. Please double check all of this, and that updated figure numbers in text match figure numbers in captions after fixing figure numbering.

Response: Thanks for your suggestion, we checked and updated the figure numbers in the revised manuscript.

Line comments:

Line 17: ’the entire study area’- suggest 'entire study area of Asia' or similar more specific wording to remind reader

Response: As suggested, we modified this sentence as “In terms of the entire study area of Asia”. Line 17

Line 21: ‘For temperature extremes…’ To prepare reader for sub-regional reporting, I suggest starting with wording such as 'In terms of sub-regional changes in temperatures extremes, the largest increases in...'

Response: As suggested, we modified this sentence as “In terms of sub-regional changes in temperatures extremes…”. Line 21

Line 45: ‘North-East of India’ - suggest 'North eastern' or 'The North East of'

Response: As suggested, the description “North-East of India” was revised as “North eastern”. Line 44

Line 63: ‘using the five Earth system models (ESMs)’ - which five? (author use of the word 'the' seems to imply these were introduced previously?)- suggest removing ‘the’ unless authors have previously introduced these models

Response: As suggested, we removed “the” in this sentence. Line 61

Line 69: ‘global warming 1.5 °C and 2 °C,’ – suggest 'global warming of 1.5 °C and 2 °C'

Response: As suggested, we modified this sentence as “In eastern Asia, under global warming of 1.5 °C and 2 °C…”. Line 68

Line 70: ‘Plateau and when’ – there are two complete phrases run together with 'and'- please separate the phrases into different sentences.

Response: As suggested, we separated the phrases into different sentences as follows:

“In eastern Asia, under global warming of 1.5 °C and 2 °C, China is expected to grow at a faster rate than the global mean, and there is a strong warming in the Tibetan Plateau (You et al. 2020). When

studying changes in local climate between 1.5 °C and 2 °C of global warming, non-GHGs influences need to be considered (King et al. 2018; You et al. 2020).” Lines 68-71

Line 90: ‘on human body’ - suggest 'to human health and well-being' or similar wording

Response: As suggested, we replaced “on human body” with “to human health and well-being”. Line 89

Line 93: ‘is lacked’ – suggest ‘is lacking’

Response: As suggested, we replaced “lacked” with “lacking”. Line 92

Lines 76-93: thanks to the authors for adding in more detail about previous work, how this work represents a significant advance- I was much more convinced of the novelty/importance of this work after reading this paragraph

Response: Thanks for your positive evaluation.

Line 95: ‘increasing’ – suggest ‘will increase’ or better 'will increasingly become a threat to...'

Response: As suggested, we replaced “increasing” with “will increasingly become a threat to…”. Line 94

Line 107: ‘have become more’ - can authors be more specific here about what is better in these ESMs that make it worthwhile to do the analysis? they suggest/hint at this earlier but would be helpful to be more explicit here.

Response: Thanks for your suggestion, we clarified as follows:

“With continuous development, some models have evolved from the original climate system model to more complete ESMs that considers the coupling of Earth system processes (including atmosphere, ocean, land, sea and ice, and biogeochemical cycles) and can provide a more realistic and complete analysis of the whole Earth-atmosphere system.” Lines 105-107

Line 113: ‘in eastern and southern Asia using’ – suggest adding in: 'Asia (here defined as 0-60°N, 70-150°E.) using' and then eliminating the sentence at the end of the paragraph because the last

sentence doesn't seem to fit, and this information can be included in this sentence here.

Response: As suggested, we removed the last sentence and modified this sentence as follows:

"…in eastern and southern Asia (here defined as 0-60°N, 70-150°E) using…" Line 112

Lines 146-147: 'thus, we directly used the CMIP6 multi-model ensemble (MME) mean' - please specify that the authors are referring to the sub-selection of models here so the reader is not confused thinking authors are using the 'entire' ensemble of ~40-50 CMIP6 models

Response: Thanks for your suggestion, we have mentioned that there are seven models carrying out both SSP3-7.0 and SSP3-7.0-lowNTCF experiments in Lines 138-139. Here we emphasized once again and modified this sentence as follows:

"At present, there are few models carrying out both SSP3-7.0 and SSP3-7.0-lowNTCF experiments as mentioned above, thus, we directly used the seven-model ensemble mean…" Lines 145-146

Line 151: 'high level of resolution' - this is subjective- I suggest removing and simply describing the resolution.

Response: As suggested, we modified this sentence as follows:

"It spans the period from 1979 to the present with a resolution of 0.5° × 0.5°." Line 151

Line 182: 'were are' – typographic error/fix wording

Response: Corrected. Line 179

Lines 209-210: 'Regionally, the greatest warming of more than 1.5 K is found in central and northern Asia and in northern North America, particularly in the Arctic, where warming is greater than 2.5 K.' - might be helpful to cite a paper here that has shown this previously as I'm sure Arctic amplification of warming is not a new finding here

Response: Thanks for your suggestion, we discussed and cited papers here.

"Previous studies have also shown that the Arctic warms at a much higher rate than the rest of the globe in recent decades (Screen and Simmonds, 2010; Pithan and Mauritsen, 2014)" Line 212-214

Lines 216-217: 'The reduction of non-methane SLCF emissions results in an average increase of 0.19

K in global mean SAT in the MME results, ranging from 0.06 K to 0.29 K across different models (Table 3).' – thanks to authors for including the spread in model results to give the reader a sense of the uncertainty- please do so elsewhere where reporting mean model results.

Response: Thanks for your positive evaluation, we added uncertainty/spread to all multi-model mean in the revised manuscript.

Lines 227-233: Where is the model spread/range? Only one number is reported ("27.7%"- this makes the answer sound so certain, but I'm sure there's disagreement in the models in terms of magnitude, correct?)

Response: As suggested, we added uncertainty/spread as follows:

"By 2050, the hottest day and warm days increase by 2.2 ± 0.7 K and 27.7 ± 13.1 % under the SSP3-7.0 scenario and by 2.6 ± 0.6 K and 34.9 ± 13.3 % under the SSP3-7.0-lowNTCF scenario, respectively. Compared to the hottest day and warm days, larger increases occur in the coldest day and warm nights, which increase by 2.8 ± 0.8 K and 35.9 ± 14.9 % under the SSP3-7.0 scenario and 3.2 ± 1.0 K and 43.1 ± 15.6 % under the SSP3-7.0-lowNTCF scenario, respectively." Lines 230-233

Lines 239-247: In my opinion, this description of previous work and comparison with results from other regions outside the study region seems to fit more in the Introduction or Discussion

Response: Thanks for your suggestion, we compared with results from other regions in the Discussion as follows:

"Such strong local effects of SLCFs on temperature extremes were also revealed in other high emission and population density regions such as Europe and the United States (Sillmann et al., 2013a; Samset et al., 2018; Luo et al., 2020)." Line 415-417

Lines 256-257: 'Concomitant with the increases in TNn and TN90p, TR is also projected to increase over the entire study area in the future.' As far as I can determine from looking at the maps, this is not true- there are blue areas on the map, indicating a decrease in days with tropical nights on the Tibetan Plateau, unless I am misinterpreting the map and colorbar- either way, TR does not consistently increase over the entire study area.

Response: Thanks for your suggestion, we clarified this sentence as follows:

"Concomitant with the increases in the coldest day and warm night, tropical night is also projected to increase over the most entire study area in the future." Lines 256-257

Lines 281-284: please include the model spread here, not just the mean

Response: As suggested, we added model spread as follows:

"The domain-averaged values of the hottest day, warm days, tropical night, and warm spell duration across the entire study area have increases of 1.4 ± 0.4 K, 37.4 ± 11.3 %, 8.9 ± 2.3 days, and 9.1 ± 2.0 days under the SSP3-7.0 scenario, respectively (Fig. 8). Considering the effects of non-methane SLCF reductions, these extreme temperature indices show increases of 1.7 ± 0.4 K, 42.2 ± 11.0 %, 10.6 ± 2.8 days, and 10.1 ± 1.7 days under the SSP3-7.0-lowNTCF scenario, respectively. Future reductions in non-methane SLCFs result in additional increases of 0.3 ± 0.1 K in the hottest day, 4.8 ± 2.2 % in warm days, 1.7 ± 0.8 days in tropical night, and 1.0 ± 0.4 days in warm spell duration for the entire study area. The hottest day, warm days, and warm spell duration all show their largest regional increases in NIN and NC, reaching 0.5 ± 0.2 K, 8.5 ± 2.9 %, and 2.8 ± 1.1 days and 0.4 ± 0.2 K, 5.9 ± 2.7 %, and 3.0 ± 1.3 days, respectively. The models agree on the sign of the change in tropical night across regions, and the largest increases occur in NC and the SCB, at 5.1 ± 2.5 and 4.9 ± 3.3 days, respectively" Line 282-290

Lines 284-285: 'In general, the SCB is most strongly affected by extreme temperatures under both scenarios' - based on which metric? looking forward two sentences, the authors indicate that the other regions are most strongly affected. This wording seems to directly contradict what the authors say later in the paragraph.

Response: Thanks for your suggestion, we removed this sentence in the revised manuscript.

Lines 293-294: 'Figure 7 shows the time series of changes in annual mean precipitation indices averaged across the entire study area under the SSP3-7.0 and SSP3-7.0-lowNTCF scenarios from 2015 to 2050 relative to the reference period' - can authors please show range of difference in gray shading around the grey line? Or does this make the figure too complicated?

Response: Thanks for your suggestion, we omitted the grey lines in Figure 1, Figure 6 and Figure 9 in order not to complicate the figures.

Lines 296-300- here and elsewhere, please include range in model results, not just the mean/median

Response: As suggested, we added model spread as follows:

"By 2050, the values of heavy precipitation days, the maximum consecutive 5-day precipitation, and total wet-day precipitation have increased by 1.6 ± 0.6 days, 2.7 ± 1.0 days, and 46.7 ± 16.4 mm under the SSP3-7.0 scenario, respectively. The increases in precipitation decrease drought events, resulting in a reduction in maximum consecutive dry days of 2.4 ± 1.5 days" Lines 298-301

Lines 304-306: 'Under the SSP3-7.0 scenario, R10 decreases by approximately 2 days in central India, the Indo-China Peninsula, and by more than 4 days in the southeastern Qinghai-Tibet Plateau, SWC, and parts of Indonesia, while it increases in all other regions of the entire study area (Fig. 8a).' - do the models even agree on the sign of change in all these locations? How confident are the authors in projected changes in local precipitation if the models don't even agree on the sign of change?

Response: Thanks for your suggestion, we clarified as follows:

"…However, it is worth noting that the consistency among models is poor." Lines 309-310

Lines 310-311: 'significant'- how is this significant? Based on what metric? Most of the boxplots in Figure 9b show nearly complete overlap in the 25th-75th percentile estimates.

If the authors are going to suggest that the results are 'significantly different' can they at least use a significance test, such as a Mann-Whitney test to determine if the medians/distributions are statistically distinguishable?

Response: Thanks for your suggestion, we removed the word 'significant'.

Line 320: 'significant increase'- see earlier comments on significance

Response: Thanks for your suggestion, we removed the word 'significant'.

Lines 331-334: again, how is significance determined?

Response: Thanks for your suggestion, we removed the word 'significant'.

Line 337: 'However, CDD increases in NIN, the southeastern Tibetan Plateau, and the southern Yangtze River (Fig. 8a and b).' – do the models even agree on the sign of change across all of these

locations, let alone the magnitude? I don't see stippling on the maps in many of these areas.

Response: Thanks for your suggestion, we modified this sentence as follows:

"However, consecutive dry days increases in NIN, the southeastern Tibetan Plateau, and the southern Yangtze River (Fig. 10j and k) with low inter-model agreement" Line 344-345

Lines 338-341: 'Future reductions…'- as far as I can determine, this is the first time that sign agreement has really been acknowledged for precipitation changes- I think the significance and sign agreement needs to be better addressed elsewhere, not just here.

Response: Accepted. We calculated the signal-to-noise ratios (please see the supplement) and added the description of the signal-to-noise ratios and sign agreement of the multi-model ensemble results to elaborate the confidence of the results in the text. Please see the revised manuscript.

Lines 343-344: 'Notably, the increases in CDD in India and eastern China are accompanied by increases in the frequency and intensity of extreme precipitation, which may be related to the probability distribution of future precipitation.' – see previous comments about model agreement on sign of change, significance considerations.

Response: Thanks for your suggestion, we added a description of the sign consistency of the results in the revised manuscript, please see previous response.

Lines 349-351: 'In general, along with the increase in average temperature, future reductions in the emissions of non-methane SLCFs will increase the intensity and frequency of extreme precipitation and decrease the occurrence of extreme droughts in the entire study area.' This and other similar statements seem to be worded as incredibly certain, but don't seem to be backed up by clear evidence from my reading of the figures. For example, when I look at this figure, it looks like this is somewhat true in southern China for R10 and RX5day, and R95p, but not true for CDD, which shows mixed results, and very little model agreement on sign of change.

Response: Thanks for your suggestion, we removed the relevant description and added some descriptions of the sign consistency and the signal-to-noise ratios of the results in the revised manuscript.

Lines 355-356: ‘along with a decrease in CDD in northern China’ - I see some light blue in Figure 8 with some stippling in isolated areas in eastern China, but a lot of China shows no stippling, which gives me little confidence in the sign, let alone magnitude, of these projections, unless I am misinterpreting the results, if so, please explain.

Response: Thanks for your suggestion, we removed the relevant description

Lines 357-359: ‘Consequently, our results indicate that future non-methane SLCF reductions may alleviate the observed precipitation anomaly pattern of “southern flood/northern drought” over eastern China, but may not have apparent mitigating effects on the precipitation reduction in India.’ – I am not familiar with the southern flood/northern drought research, but when I look at this figure, I see a mostly east/west, not north/south pattern here with little sign agreement, so I don't know what evidence this statement is based on.

Response: Thanks for your suggestion, we removed the relevant description

Lines 390-391: ‘The climate factor is also the largest contributor to the increase in population exposure, followed by the climate- population interaction factor and the population factor in NC, the SCB, and SC’ - please be more specific- to precipitation extremes?

Response: Thanks for your suggestion, we clarified as follows:

“For population exposure under extreme precipitation, the climate factor is also the largest contributor to the increase in population exposure, followed by the climate-population interaction factor and the population factor in NC, the SCB, and SC (Fig. 14b-d).” Line 390-392

Lines 401-403: ‘Our results show that non-methane SLCF reductions will exacerbate the warming effect caused by GHGs, resulting in increases in extreme temperature and precipitation events compared to the standard SSP3-7.0 scenario.’ - if reduced alone, correct?

Response: Yes, we modified this sentence as follows:

“Our results show that non-methane SLCF reductions alone will exacerbate the warming effect caused by GHGs…” Lines 402-403

Lines 420-421: ‘Notably, large differences in extreme precipitation changes are found at the regional scale in response to reductions of non-methane SLCFs and only aerosols, with changes in opposite

directions observed in some regions' – yes, I appreciate the authors noted this, and I think this should be noted when individual results are presented if the models don't consistently even agree on sign of change (in the Results section).

Response: Thanks for your suggestion, we noted this in the Results section as follows:

"For heavy precipitation days, the maximum consecutive 5-day precipitation, and total wet-day precipitation, there is a robust increase in the southeastern Qinghai-Tibetan Plateau, SWC and the middle and lower reaches of the Yangtze River, with relatively good inter-model agreement and high confidence. In SWC, these extreme precipitation indices increase by 2.5 ± 1.9 days, 2.5 ± 1.5 mm, and 37.5 ± 22.6 mm, and these increases are greater or comparable to the effects of GHGs. By contrast, consecutive dry days decreases by 1.4 ± 1.2 days in SWC due to increases in the frequency and intensity of heavy precipitation events. Notably, there are still large uncertainties in the model's simulations of extreme precipitation, which may attribute to model resolution and natural variation (Li et al., 2014; Deser et al., 2012)." Lines 421-427

"Similarly, the inter-model variation in population exposure is large, which arises from differences in model simulations of extreme precipitation." Lines 433-434

Line 431: 'contrary'- opposite in sign?

Response: Yes, and we replaced "contrary" with "opposite in sign". Line 440

Lines 433-435: 'The latter is usually used as an indicator of flooding, suggesting that heavy precipitation associated with natural disasters will be aggravated in the future due to non-methane SLCFs reduction.' True in southern/SE China, but really not much agreement on sign anywhere else if I am interpreting the maps correctly.

Response: Thanks for your suggestion, we clarified as follows:

"The latter is usually used as an indicator of flooding, suggesting that heavy precipitation associated with natural disasters will be aggravated in the future due to non-methane SLCFs reduction in certain areas, especially in SC and SWC." Lines 443-445

Line 435: what is 'well assessment'?

Response: We modified this sentence as follows:

"Secondly, population exposure is an effective indicator for climate change risk assessment." Lines 445-446

Lines 436-437: 'The reduction of non-methane SLCFs will result in the exposure of millions of people to extreme events, and up to tens of millions in densely populated areas' – This sentence gives me the impression that these extreme events wouldn't have happened in the SSP3-7.0 scenario without NTCF reductions. However, my understanding looking at the SSP3-7.0 vs lowNTCF figures is that under SSP3-7.0 extremes also occur, but the elimination of NTCF just makes them more frequent/likely/worse. Is this correct? If so, can the authors more carefully word these summary findings please?

Response: Thanks for your suggestion, we clarified as follows:

"Compared to SSP3-7.0 scenario, the reduction of non-methane SLCFs will result in additional exposure of millions of people to extreme events, and up to tens of millions in densely populated areas, such as northern India…" Line 446-448

Line 443: 'that some of the extreme indices were not well fitted' – what does this mean?

Response: The linear trend of some indices is not well fitted in Allen et al. (2020). For example, less robust results are generally observed when it comes to wettest day and CDD, with R^2 of 0.485 and 0.049. The linear trends for CDD even lack significance.

Lines 452-455: thanks to the authors for contextualizing this.

Response: Thanks for your positive evaluation.

Lines 458-461- this description seems more accurate to me than what is stated on lines 436-437 (see comment above)

Response: We clarified the sentence on lines 446-448, please see the above response.

Line 486: 'has'- have

Response: Corrected.

Lines 493-494: 'which may lead to underestimation of the impact of SLCF emissions reductions in China' - wouldn't this lead to an overestimation if simulated SLCF emissions are higher than they actually are in reality (so future emissions reductions would be less impactful because they already happened, so the models would overestimate the real impacts of SLCF reductions bc we've already experienced these impacts?)? Or am I misinterpreting the impact of lower emissions? Or what the authors intend to say?

Response: What we intended to say is that the CMIP6 database underestimate emission reductions obtained from China's Action Plan in the base year, and the emission bias in the base year would pass to the future and lead to different emission mitigation pathways (Tong et al., 2020). Therefore, future emissions reductions in China are likely to be more intense than those in CMIP6, leading to more impactful to climate change in China compared to present day. We clarified this sentence as follows: "For example, the SSP-RCP global emissions scenarios used in CMIP6 do not fully consider the rapid pollution controls enacted in China since 2013 under the Air Pollution Prevention and Control Action Plan. Consequently, CMIP6 database underestimate emission reductions obtained from China's Action Plan, which may lead to underestimation of the impact of SLCF emissions reductions in China (Wang et al., 2021; Tong et al., 2020), …" Lines 499-502

Lines 496-499: 'What policymakers, the public, or the media need to know'- I will leave this wording choice up to the authors, but in my previous review I was not intending that the authors needed to explicitly address the conclusion to policymakers and the media- instead, I was suggesting the authors contextualize their results as they have done in the Discussion/Conclusion so as not to be misinterpreted by science reporters etc. In my opinion, the authors can eliminate the phrase 'What policymakers, the public or the media need to know is that' and start the sentence with 'Air pollution…'

Response: As suggested, we modified this sentence as follows:

"Air pollution is dangerous to human health, and there is no doubt that we need clean air, but more importantly efforts to reduce GHGs need to be doubled in order to simultaneously mitigate climate change and improve air quality (Quaas et al., 2022; McKenna et al., 2021)." Lines 506-508

Suggestions for figures:

For figures with maps:

In my previous review, I suggested the authors use one colormap for showing changes through time, and another colormap for showing differences among experiments so they can be quickly/easily visually distinguished. The authors combined color bars for the climate change maps, but kept the same color maps (the same colors) to show the differences among experiments- I suggest using different color maps (a different divergent color map for showing differences vs changes through time) so the reader can easily determine that the maps are not showing the same things.

Response: As suggested, we changed the colormap for showing differences among experiments, please see figures in the revised manuscript.

For all line graphs: authors reference 'green' lines in figures, but lines are now grey- please double check figure captions and change where appropriate.

Response: We omitted the grey line in Figure 1, Figure 6 and Figure 9.

Line 775/776 (Figure 1 caption): may be helpful to add a line something like 'if red line only is shown (g,h), this indicates that ssp370 and ssp370-lowNTCF overlap" or other wording to let reader know why there is only one line in bottom panels (g,h).

Response: As suggested, we added "only red line in (g, h) indicates that ssp370 and ssp370-lowNTCF overlap." in caption of Figure 1 in the revised manuscript. Lines 792-793

Figure 2 (line 780): figure panels are so small, I can't see text, etc.

Instead of this being a figure that is 2 (columns) x 8 (rows), can the authors make this two sets of double columns (4x4) so precip is next to temperature with a little space in the middle to separate them visually? This might help fill the white space on the page and improve readability bc the maps could be larger?

Response: We added a third column "CMIP6 minus observations" in response to Referee 2 and reformatted this figure.

Figure 2 (line 785): The stippling in the maps changes from figure to figure - please consistently use stippling- are significant or insignificant regions stippled? Also, the authors changed one figure in

response to my comments, but the same comment could be made for all figures showing warming- showing where models agree on warming in warming scenarios is a low bar that doesn't show meaningful information, and obscures the coloring underneath- this is the case in not only Figure 3, which the authors changed, but also Figure 2, (for ERF changes not warming), Figure 5, Figure 8, etc.

Response: Thanks for your suggestion, we stippled on all maps where models disagree in the revised manuscript.

Figure 3: what about showing the range of differences in grey shading around the grey line as well? Or does this make the plot too complicated to interpret? Also, as noted above, please change 'green lines' to 'grey lines' in caption here and elsewhere where necessary.

Response: We have omitted the grey line in Figure 1, Figure 6 and Figure 9 in order not to make the figures more complicated.

Figure 5: I suggest adding units on colorbar label, or add in units in the figure caption as in earlier figures- the labels in the upper right are hard to find/read

Response: As suggested, we moved the location of the units. Please see Figure 7, Figure 10, and Figure 12 in the revised manuscript.

Figures 6,9,10: Why are country/political boundaries only shown in Figure 5 and 8, and not in other figures with maps (such as maps in Figs 6,9,10)? Can the authors please be more consistent with showing these borders for reader orientation unless there is a specific reason to exclude them?

Response: As suggested, we added country borders/boundaries to all the maps in the revised manuscript.

Figure 6, 7 and other graphs with y axes that cross zero: here and in other figures, a horizontal reference line marking zero on the y axis can be helpful.

Response: As suggested, we added a horizontal reference line marking zero on the y axis in Figure 6, Figure 8, Figure 9 and Figure 11 in the revised manuscript.

Figure 11, Figure 12: legend has an error- ‘climte’

Response: Corrected.

Reference:

Deser, C., Knutti, R., Solomon, S., and Phillips, A. S.: Communication of the role of natural variability in future North American climate, Nat. Clim. Chang., 2, 775–779, https://doi.org/10.1038/nclimate1562, 2012.

Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., and Wehner, M.: Changes in annual extremes of daily temperature and precipitation in CMIP6 models, J. Clim., 34, 3441–3460, https://doi.org/10.1175/JCLI-D-19-1013.1, 2021.

Li, J., Yu, R., Yuan, W., Chen, H., Sun, W., and Zhang, Y.: Precipitation over East Asia simulated by NCAR CAM5 at different horizontal resolutions, J. Adv. Model. Earth Syst., 7, 774–790, https://doi.org/https://doi.org/10.1002/2014MS000414, 2015.

Pithan, F. and Mauritsen, T.: Arctic amplification dominated by temperature feedbacks in contemporary climate models, Nat. Geosci., 7, 181–184, https://doi.org/10.1038/ngeo2071, 2014.

Samset, B. H., Sand, M., Smith, C. J., Bauer, S. E., Forster, P. M., Fuglestvedt, J. S., Osprey, S., and Schleussner, C. F.: Climate Impacts From a Removal of Anthropogenic Aerosol Emissions, Geophys. Res. Lett., 45, 1020–1029, https://doi.org/10.1002/2017GL076079, 2018.

Screen, J. A. and Simmonds, I.: The central role of diminishing sea ice in recent Arctic temperature amplification, Nature, 464, 1334–1337, https://doi.org/10.1038/nature09051, 2010.

Sillmann, J., Stjern, C. W., Myhre, G., Samset, B. H., Hodnebrog, Ø., Andrews, T., Boucher, O., Faluvegi, G., Forster, P., Kasoar, M. R., Kharin, V. V., Kirkevåg, A., Lamarque, J. F., Olivié, D. J. L., Richardson, T. B., Shindell, D., Takemura, T., Voulgarakis, A., and Zwiers, F. W.: Extreme wet and dry conditions affected differently by greenhouse gases and aerosols, npj Clim. Atmos. Sci., 2, 1–7, https://doi.org/10.1038/s41612-019-0079-3, 2019.

Tong, D., Cheng, J., Liu, Y., Yu, S., Yan, L., Hong, C., Qin, Y., Zhao, H., Zheng, Y., Geng, G., Li, M., Liu, F., Zhang, Y., Zheng, B., Clarke, L., and Zhang, Q.: Dynamic projection of anthropogenic emissions in China: methodology and 2015-2050 emission pathways under a range of socio-economic, climate policy, and pollution control scenarios, Atmos. Chem. Phys., 20, 5729–5757, https://doi.org/10.5194/acp-20-5729-2020, 2020.

Response to Referee #2:

We are grateful to the referees for their time and energy in providing helpful comments and guidance that have improved the manuscript. In this document, we describe how we have addressed the reviewer's comments. Referee comments are shown in black and author responses are shown in blue text.

Li et al compare CMIP6 simulations with/without NTCFs and show that ERF increases due to NTCF reductions. They compare various climate extremes indices over eastern and southern Asia between the two scenarios. The authors have largely responded well to the previous reviewer comments. However, I have some additional points that I think should be considered prior to publication. Line numbers refer to the tracked revised manuscript.

Response: Thank you for your positive evaluations. All the questions and concerns have been carefully answered and the paper has been revised accordingly.

Major comments

1. Are the findings statistically significant? In some places the word 'significant' is used (e.g. 'A significant increase of more than 40 days occurs…', line 257), however no details of statistical testing are given. The standard deviations overlap in figures such as fig. 4, 7, and in fig 6 the boxplots overlap. For the map plots such as fig 5, 8, dots are used to indicate that a given number of models agree on the sign. It would be more meaningful if dots represented regions where the differences due to NTCFs are not statistically significant. Having models agree on the sign is not convincing that there is a significant difference between the two scenarios. Lastly, the uncertainty ranges quoted on lines 395-397 imply that the impact of NTCF reductions does not make a statistically significant impact. I strongly recommend implementing statistical testing on the results.

Response: Thanks for your suggestion,

i) We redescribed sentences with "significant" in the revised manuscript.

ii) We stippled on all maps where models disagree in the revised manuscript and added the signal-to-noise ratio in the Supplements to describe the confidence level of the multi-model results.

iii) We also implemented statistical testing on the MME result. Taking TX90p and R95p for example, we found that the regions with significance at ≥95 % confidence level from the t-test are broadly in line with those with signal-to-noise ratio more than 1 (Figure R1, Figure R2).

Therefore, we chose the signal-to-noise ratio to quantitatively characterize the confidence of the multi-model ensemble signal in the revised manuscript.

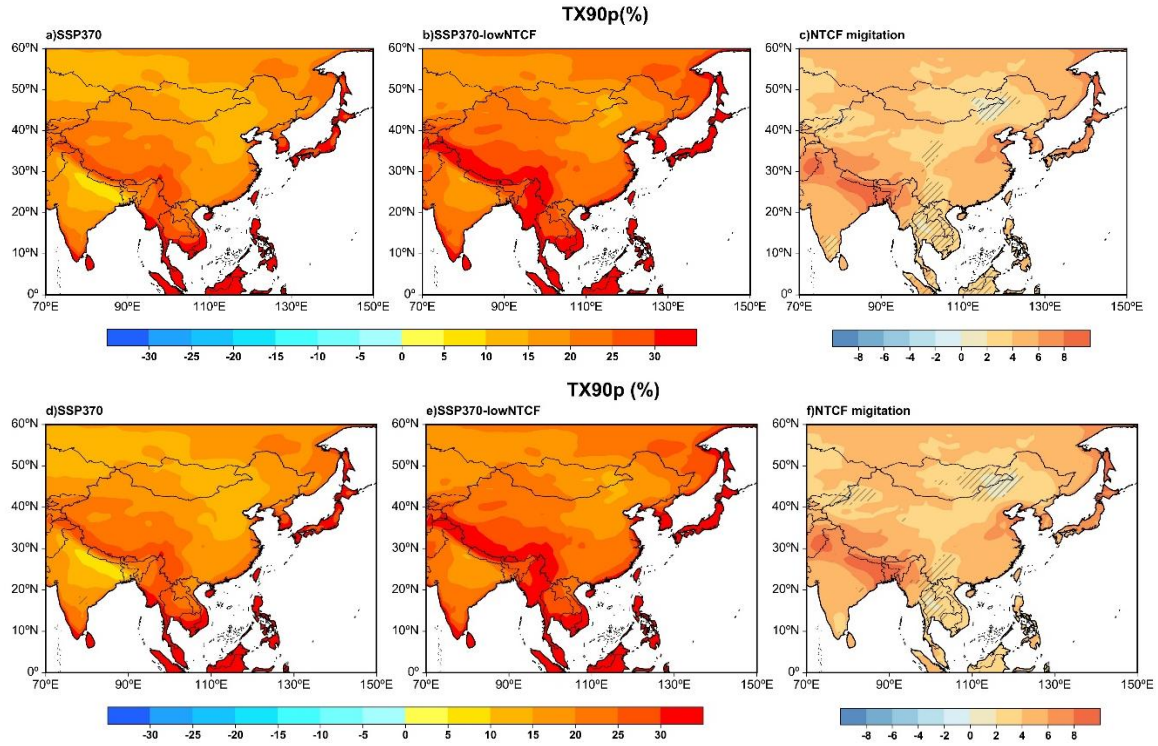

Figure R1 Spatial patterns of changes in warm days (TX90p) during 2031-2050 in Asia under the SSP3-7.0 (left column) and SSP3-7.0-lowNTCF (middle column) scenarios relative to 1995-2014. The right column represents changes caused by the non-methane SLCFs mitigation. The slipped regions in (a~c) and (d~f) indicate that the signal-to-noise ratio is less than 1 and significance at <95 % confidence level from the t-test, respectively.

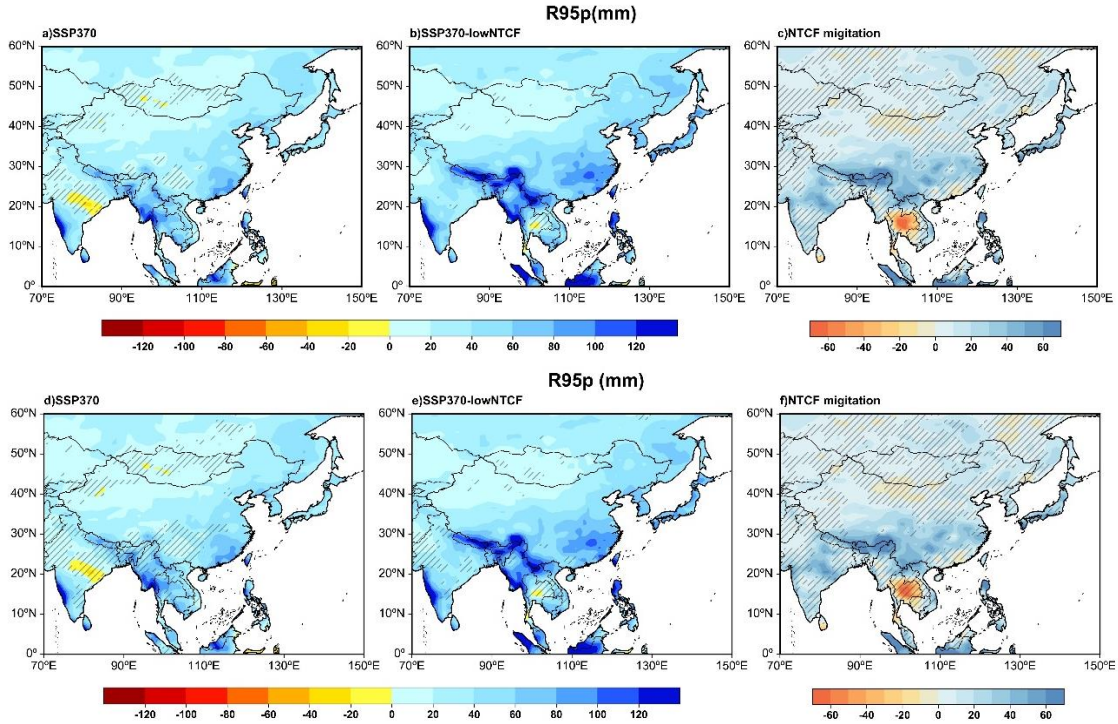

Figure R2 Spatial patterns of changes in total wet-day precipitation (R95p) during 2031-2050 in Asia under the SSP3-7.0 (left column) and SSP3-7.0-lowNTCF (middle column) scenarios relative to 1995-2014. The right column represents changes caused by the non-methane SLCFs mitigation. The slipped regions in (a~c) and (d~f) indicate that the signal-to-noise ratio is less than 1 and significance at <95 % confidence level from the t-test, respectively.

2. I am uncomfortable with the framing that NTCF reductions lead to increased ERF; this is a blanket statement and not true of all possible NTCF mitigation scenarios. At worst, it could lead readers to believe that it is not worth pursuing NTCF emissions reductions because of the extra warming it would incur. As the authors are aware, not all NTCFs behave the same way; tropospheric ozone and black carbon both have warming effects. Indeed, mitigating these emissions has been recommended as a way to complement CO2 reduction measures and improve air quality (see e.g. the UNEP Integrated Assessment of Black Carbon and Tropospheric Ozone, 2011). Clearly it is reductions in non-BC aerosols that are driving the increase in ERF in the SSP3-7.0-lowNTCF scenario; this point needs to be made more clearly. For example, line 222: “the cooling effect of ozone reduction is significantly weaker than the warming effect of aerosol reduction.” This is true for the specific scenario analysed here, but would not necessarily be true for alternative low-NTCF scenarios.

Similarly, line 273: “Overall, the combined reductions of aerosols, ozone, and their precursors have an additional warming effect on Earth’s climate system, exacerbating the surface warming and extreme temperature events caused by GHGs.” Ozone reductions cannot contribute to the additional warming; the additional warming in the low-NTCF scenario can only come from aerosol reductions. More care should be taken throughout the paper to caveat the results appropriately.

Response: Thanks for your suggestion,

i) We have mentioned in the discussion that the important impact of SLCFs on climate change in the short term does not mean that reducing air pollutants is harmful to the climate, but rather that it generates additional warming to the climate, amplifying the temperature and extreme precipitation caused by GHGs changes. Please see Lines 467-480 for more details.

ii) As suggested, we caveated the results more appropriately and clarified as follows:

“Although the reduction of SLCFs leads to a decrease in the aerosol optical depth (AOD) on a global scale (Fig. S1), not all SLCFs behave the same way and it is reductions in non-BC aerosols that are driving the increase in ERF in the SSP3-7.0-lowNTCF scenario.” Lines 201-203

“In general, under the SSP3-7.0-lowNTCF scenario, the combined reductions of aerosols, ozone, and their precursors causes further warming, suggesting that the cooling effect of ozone reduction is somewhat weaker than the warming effect of aerosol reduction.” Lines 223-225

“Overall, although ozone reductions cause cooling, the combined reductions of aerosols, ozone, and their precursors have an additional warming effect on Earth’s climate system, exacerbating the surface warming and extreme temperature events caused by GHGs.” Lines 273-275

3. Related to the above point, it would be useful to understand how NTCFs differ spatially over the study region during 2031-2050. Could maps of ground-level ozone concentration and aerosol optical depth be included?

Response: As suggested, we added maps of ground-level ozone concentration and aerosol optical depth in Figure S1 in Supplements.

4. A small number of models performed the SSP3-7.0 and SSP3-7.0-lowNTCF simulations, and the

authors analyse the multi-model mean. How skewed are the model responses? Would you see large differences in the results if you analyse the multi-model median? I am aware, for example, that the BCC-ESM1 model (one of the models analysed) contains large biases in aerosol optical depth relative to other models in certain regions, but do not know if these correlate with biases in the climate extremes examined over southeast Asia.

Response: Thanks for your suggestion,

i) We mentioned that there are large differences among models related to extreme precipitation changes in the revised manuscript.

ii) There are slight differences when analyzing the multi-model median and multi-model mean under the SSP3-7.0 and SSP3-7.0-lowNTCF scenarios, but such differences are small when it comes to the impact of non-methane SLCFs reductions. As shown in Figure 8 and Figure 11, the black center lines and dots almost overlap.

iii) We analyzed the different models of the spatial distribution of AOD changes due to emission reductions of non-methane SLCFs in eastern and southern Asia. In general, the results of the BCC-ESM1 model and other models were in relatively good agreement (Figure R3).

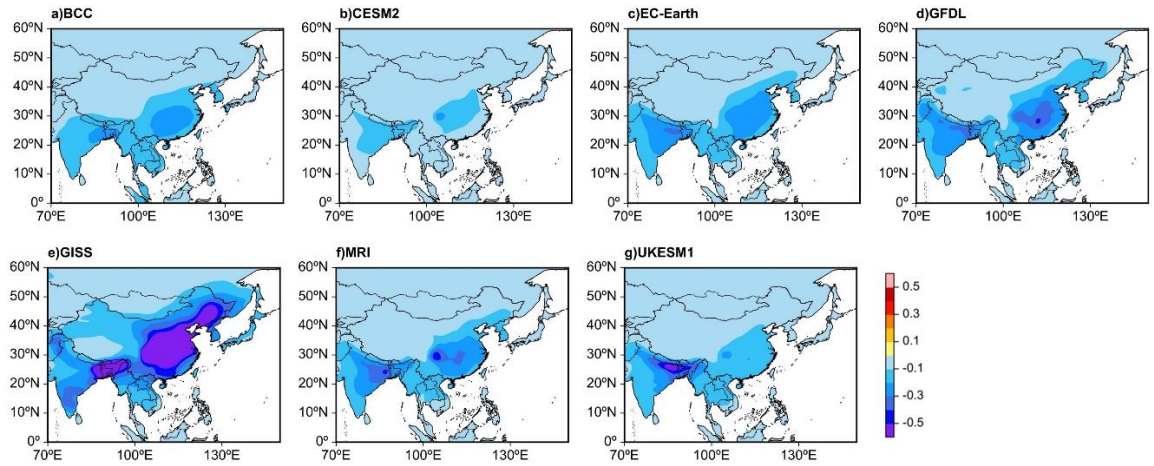

Figure R3 Spatial patterns of changes in ambient aerosol optical depth at 550 nm in Asia during 2031-2050 caused by non-methane SLCFs mitigation for different models.

Minor comments

Line 17: “in terms of the entire study area” note that the study area has not been defined by this point.

Response: Thanks for your suggestion, we modified this sentence as “In terms of the entire study

area of Asia". Line 17

Line 122: this is unclear wording - why should a lack of climate policies related to GHG reductions and air pollution controls mean that SSP3-7.0 is used as a reference scenario? In reality, SSP3-7.0 and SSP3-7.0-lowNTCF are the only pair of simulations available through CMIP6 for which the effects of NTCFs can be assessed in a controlled manner, so I would simply state this.

Response: As suggested, we revised this sentence as follows:

"The SSP3-7.0 scenario includes no climate policies related to GHG reductions and weak air pollution controls (~7.0 W m^{-2} at 2100). In this scenario, the future climate change is approximatively influenced by GHG forcing only. The SSP3-7.0-lowNTCF scenario uses the same GHG forcing path but with strong air pollution controls, and thus future climate change in this scenario is influenced by a combination of GHG forcing and reduction of non-methane SLCF emissions." Lines 121-125

Table 1: did all models provide fSST simulations for calculating ERF? And, did they provide the same number of fSST realisations as stated in Table 1?

Response: Yes, all models used in this study provide fSST simulations for calculating ERF and they provide the same number of fSST realizations. We mentioned this in the revised manuscript:

"All models used here provided SST simulations for calculating ERF and they provided the same number of realizations as stated in Table 1" Lines 141-143

Table 3: it would be useful to also report the SAT for the study area. Also note the discrepancy between the caption (SAT) and table header (Tas).

Response: Corrected.

The figure numbers and captions need to be updated; there are two figure 2's, for example. And the captions sometimes refer to green dashed lines, which are now grey (e.g. figure 1).

Response: Thanks for your suggestion, we checked and updated the figure numbers in the revised manuscript.

Figure 1: the grey dashed lines showing the difference between the two scenarios don't really make sense when reading the y axis labels. I got confused and started wondering how negative emissions are possible. I would be tempted to omit the line showing differences in this figure.

Response: As suggested, we omitted the grey line in Figure 1 in the revised manuscript.

Figure 2: great to see the evaluation work done as recommended by previous reviewers. Could I suggest adding a third column: CMIP6 minus observations?

Response: As suggested, we added a third column in the revised manuscript.

Figures 11 and 12 have spelling errors in the legend ('climte'); also please define the acronyms in the headers.

Response: Thanks for your suggestion,

i) Corrected. ii) We defined the acronyms in the headers, please see Figures 13 and 14 in the revised manuscript.

Line 168: representative of what?

Response: Thanks for your suggestion, we clarified this sentence as follows:

"In this study, we focused on the changes in the future (2031–2050) relative to the reference period (1995–2014), the results of CMIP6 MME for future extreme climate projections can be considered representative." Lines 168-169

Section 3.2 and 3.3: I couldn't review these sections properly as there were just too many acronyms that I am unfamiliar with (I am not an expert in climate extremes). Table 2 is a useful addition to the paper, but it is tiring to have to keep cross-referencing it. I suggest keeping table 2, but writing out the acronyms wherever possible so as not to interrupt the flow of text.

Response: As suggested, we write out the acronyms as much as possible in the text in the revised manuscript.

Line 331-332: you don't need the minus signs (e.g. 'significant decreases … <-50 mm')

Response: Corrected.