Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2022-422-RC1, 2022

Comment on acp-2022-422

Anonymous Referee #1

Referee comment on "Impacts of reductions in non-methane short-lived climate forcers on future climate extremes and the resulting population exposure risks in eastern and southern Asia" by Yingfang Li et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-422-RC1, 2022

Review of: Impacts of reductions in non-methane short-lived climate forcers on future climate extremes and the resulting population exposure risks in Asia

Summary:

Li et al. use CMIP6 simulations (SSP3-7.0 and SSP3-lowNTCF, part of AerChemMIP) to compare the impacts of increasing emissions of greenhouse gases+aerosols (SSP3-7.0) as opposed to reduced emissions of aerosols (Near-Term Climate Forcers, or NTCFs) and continued emissions of greenhouse gases (SSP3-lowNTCF, which is an idealized simulation that implements future air quality standards that do not reduce GHG emissions). The authors first compare global impacts of SSP3-7.0, SSP3-lowNTCF, and the difference between the two experiments (under the assumption that studying the difference isolates the climate impact of reduced non-methane NTCFs), then specifically South/East Asia. The authors report changes in Effective Radiative Forcing (ERF) and increased climate extremes. I found this paper to be generally well written and the text easy to follow, and the results are interesting and potentially important if policymakers implement short-term air pollution controls only (without reducing any greenhouse gas emissions). My main concerns are outlined below, followed by specific line comments.

Before I continue my review, I should note that I am relatively new to the research area of NTCF simulations, but I have more experience in the area of population exposure, CMIP ensemble analysis, and climate impacts of emissions. I am not a statistician or climate extremes expert, so I have focused my comments on the overall scientific quality of the paper, novelty of the work, clarity of the writing/figures, and overall climate impacts assessed.

Main Concerns:

- I am less familiar with the literature related to AerChemMIP, so I went back to re-read a few of the papers Li et al. cite in the introduction. Specifically, I read Allen et al. (2020), who conducted a very similar analysis (compared SSP3-7.0 to SSP3-lowNTCF simulations) and parts of the IPCC AR6 WG1 Chapter 6.
  My main concern is that many of the results presented here are either quantitatively or qualitatively similar to those in Allen et al.: Allen et al. compare CMIP6 SSP3-7.0 and SSP3-lowNTCF, and already make the main points this manuscript makes- specifically, climate extremes (temperature, precipitation) are intensified with reduced aerosol forcing, and this change in extremes is strongly felt over parts of southern and eastern Asia. Some (but not all) of the climate extreme indices are even already presented in Allen et al.
  I do recognize that there is some new material here because Li et al. assess a wider range of climate extreme indices, examine population-weighted climate impacts, and focus on Asia. However, I didn't notice any outstanding new findings, but instead more details on findings similar to those already published. The authors simply acknowledge that Allen et al. examine 'trends', but don't really acknowledge very much beyond that, and even re-present similar information (e.g., changes in ERF), but using a slightly different method (change in means over two time periods instead of trends, etc).

- A critique of how this experiment (SSP3-lowNTCF) is analyzed and presented here: there seems to be a history of 'idealized emissions reduction modeling' experiments that have led to misconceptions related to the impacts of aerosol reductions in the scientific literature and in the public/media (e.g., see Shindell and Smith, 2019, Nature, who argue against the realism of immediate 'zero emissions' and the associated spike in warming).
  Admittedly, SSP3-lowNTCF is a more 'realistic' scenario that doesn't implement immediate 'zero emissions' – However, I went back to the IPCC AR6 WG1 Chapter 6.6 and 6.7, and the IPCC specifically states that SSP3-lowNTCF is 'an idealized simulation of a very ambitious air-quality policy where the maximum technical potential of existing end-of-pipe technologies is explored'. How realistic is it that nations will only implement air quality, and absolutely no GHG reductions? And how are the results presented here to be interpreted (esp. by policymakers, the public, and/or the media)? I understand that there is scientific value in distinguishing among GHG vs aerosol reductions, but how is this information to be used, and what is the context?
  I bring these issues up because I think the Discussion/Conclusion could be strengthened by at least a qualitative comparison with SSP1 (which in my understanding is where the reduced aerosol emissions come from in the SSP3-lowNTCF experiments)- what are the benefits of reducing both GHG+NTCF emissions instead of reduced NTCF emissions alone? Without context, the results presented here make me think 'Air pollution reductions are harmful- they will worsen climate extremes'- Is this the takeaway the authors intend?
  IPCC AR6 Chapter 6.7 ends with some discussion of the contextualization of these

idealized results in comparison with SSP1. I suggest that both in the introduction and conclusion, the authors add wording and some discussion of the idealized nature of the experiment analyzed, and how simultaneous reductions of GHG+NTCF produce less extreme impacts on climate (see for example, Shindell and Smith, 2019, Nature, or how Allen et al. also contextualize the changes in extremes with the reduced air pollution exposure, etc.).

- Clarity of figures:
  - In several figures showing time series, there are no y axis labels or x axis labels (e.g., Figure 4, others). Also, there are acronyms in the figure subplot titles that are not defined in the figure caption- it is difficult to find what these are without searching through the text/tables, so please define all acronyms in the caption. I understand the extremes are defined in a table, but it would be helpful to not have to flip back and forth to determine what they are.
  - In many of the maps showing changes, the same colormap is used to show changes over time and differences among experiments in the same figure, which is visually confusing.
  - Several figures showing time series include red and green lines, which will be indistinguishable for a red/green colorblind reader.
  - Several of the figures with maps show dots/stippling at locations where >60% of models 'agree on sign of change'. I don't find stippling 'significant' locations in this case to be particularly helpful visually because the dots cover almost all of the map (yes, CMIP6 models show the globe warms under SSP3- why is this stippled when this is my default expectation, and has been reported before?). Perhaps more importantly, the stippling obscures the colors underneath so the reader cannot easily interpret the colors. Can the authors stipple throughout where there is disagreement?
  - Showing that CMIP6 models agree on sign of change for global warming is a frankly low bar/not a robust metric, as is the low >60% agreement threshold. Can the authors choose a more robust method (e.g., agree on magnitude of change, using something like coefficient of variation- e.g., Buzan and Huber, 2020: https://www.annualreviews.org/doi/abs/10.1146/annurev-earth-053018-060100), or at least where 75% or 90% or 95% of the models agree on sign?

- CMIP6 model validation in terms of climate extremes: I was surprised to see an evaluation of CMIP6 model results without any (as far as I could tell) mention/citation of an assessment of model performance- how well do these models simulate temperature + precipitation extremes if we are to rely of their projections of future extremes?

- CMIP6 multi-model ensemble averaging: From my reading of the paper, the authors seem to have just averaged across an ensemble of CMIP6 models- should the models be selected/weighted according to performance, or perhaps according to independence (see for example Brunner et al., 2020: https://esd.copernicus.org/articles/11/995/2020/ who suggest that equally weighting

CMIP6 models does not produce the same results as weighting them based on independence/performance)? Or are there too few models? Some mention/justification of simple ensemble averaging could be helpful.

- The authors have an ensemble of model simulations, so they should have a range of results in terms of changes in climate extremes, but the main results presented in the text and Abstract are presented as one number (e.g. 'regional average temperature on the hottest days (TXx) by 0.3 K)'- isn't there some range/spread in the results? This range is shown in the figures and mentioned occasionally in the text, but this nuance/uncertainty does not come out in the Abstract or in much of the results, where one mean/median number, with no range/uncertainty is presented. It would seem to me that the range or 1 or 2 sigma should be presented next to all of the mean/median results.

Line/Specific Comments:

Lines 11-15: 'Stringent…climate'- the authors mention that SLCF emissions reductions have been implemented, then in the next sentence state that they examine future impacts. The naïve reader could assume that the SSP3-lowNTCF simulations are realistic and are simply a continuation of past air quality/climate policy. Is this the case?

Also, this sentence conveys to me the idea that emissions reductions have already happened, but the next sentence claims to study future emissions reductions. I think some clarification could be helpful here.

Line 13: 'in Asia'- the boxes in the main text seem to be a sub-section of eastern/southern Asia, not all of Asia.

Line 14: 'SSP' – please define this acronym (authors define all other acronyms in Abstract, why not this one?).

Lines 16-18: 'The MME results…': The authors report a change in ERF that seems to be half the magnitude of that presented in the Allen et al. paper cited in the Introduction (0.44 W/m2 over the entire time period in Allen et al.). Why the difference? And also, most of the paper is not about ERF, but climate extremes, so why is this part of the focus

in the Abstract? Or is this necessary to explain why changes are observed? I don't understand why the authors need to re-report changes in ERF that were recently published unless they show something significantly new/different.

Line 16: 'Regional average temperature'- please remind reader which region (transitioned from global ERF in previous sentence)

Line 24: 'predicted'- here and elsewhere in the manuscript, isn't 'projected' the preferred term for changes in climate based on changes in boundary conditions/forcing?

Lines 58-59: 'on future climate change has been limited to the effect of aerosol forcing associated with incomplete interactive tropospheric chemistry schemes in global climate models' - except in Allen et al., who used CMIP6 ESMs, like in this study.

Lines 74-75: 'focused on trends of climate variables, and its assessment of regional climate changes, particularly climate extremes, was insufficient.' – this seems subjective- the extreme indices they assessed showed qualitatively similar results to the ones presented here. The time series presented in Allen et al. also show similar information, so what makes this 'insufficient'?

Line 76: Yes, the IPCC AR6 does present mostly a global overview in terms of temperature impacts (figure 6.23 in AR6 WG1), but Allen et al. already report that impacts are most intense in parts of Asia.

Lines 137-139: 'Gridded population datasets for 2000 and 2040 under SSP3 were used to represent the population during the reference and future periods, respectively.' – which dataset? GPWv4 and SSP projections? Citation?

Figure 2 and other figures with maps: see main comment about stippling significance, which covers information/colors presented on maps. Also, please use separate colormap to show differences in time periods vs differences among experiments to visually distinguish, unless there is a specific reason to use the same colormap in both.

Lines 145-149: I am more familiar with attempting to distinguish among causes of impacts (changing climate or changing population) by assessing changes in population-weighted impacts using static (present) population, and comparing to results using dynamic population. Does this produce the same result as the equation/method here?

Line 180 ‘3.2 Changes in temperature extremes in Asia’ – I was a bit surprised not to notice a mention of how well these models simulate observed frequency etc of past, observed extremes. The authors jump straight into projected extremes- are there papers showing that these temperature and precipitation extreme indices are well captured by the climate models used here? See for example (Li et al.: https://journals.ametsoc.org/view/journals/clim/34/9/JCLI-D-19-1013.1.xml, Kim et al., 2020: https://www.sciencedirect.com/science/article/pii/S2212094719302439; Yang et al., 2021: https://link.springer.com/article/10.1007/s00376-021-0351-4).

Related to this, Yang et al. (above) report that not all models perform equally well in terms of simulation of temperature extremes over China, and choose to analyze projections from a sub-selection of models that better simulate observed climate- have the authors considered doing this?

Figure 4: where is the region that the average encompasses? I don’t think the region boxes are shown until Figure 6. How are these sub-regional boxes chosen? Are these SREX regions? Or did the authors just choose boxes to maximize signal after they ran the analysis? Also, the regional maps (like Figure 5) could be easier to interpret if the authors include country borders.

Figure 7: Here and previous figure showing time series: please do not include red and green for red-green colorblind readers (why not red, blue, grey?).

Line 245: ‘the extreme precipitation changes are more significant’ – how is significance determined? Do the authors mean ‘larger’ or something similar?

Line 250: Here and Figure 8: drawing boxes outlining the regions of interest on the anomaly maps would be helpful to determine where spatially the location is that the authors are discussing in the text.

Line 293: ‘future precipitation distribution in SC could be more heterogeneous under high SSP scenarios.’ Spatial distribution or PDF distribution? Please be more specific, as I wasn’t sure what a ‘heterogeneous distribution is’ if it’s a histogram/PDF.

Figures 11/12: Typo in figure ‘climte’

Lines 329-333: ‘climate factors under both the SSP3-7.0 and SSP3-7.0-lowNTCF scenarios in the four selected regions (Fig. 11). This result suggests that climate change is the primary driver of population exposure to extreme temperature events, followed by changes in the climate population interaction factor, with population change contributing

the least.'

I am a bit confused about the wording here- how are we to distinguish SLCF from climate? I think a clarifying phrase and breaking up some long sentences could be helpful as authors explain.

Lines 346-347: 'in increases in extreme temperature and precipitation events' – this is relative to the background increases already experienced in SSP3-7.0, correct? Please clarify.

Lines 368-378: See my main comment above: can the authors benchmark/qualitatively compare previously published findings of reductions of both GHG+NTCF under SSP1-2.6 or a similar experiment? What are the benefits of reducing both vs just reducing NTCFs, or the combined impacts? And I think a further discussion/emphasis of the realism of NTCF reductions only would be helpful- are countries going to only reduce SLCFs? Have they already? How realistic are these results in a real-world context?