nature portfolio

Peer Review File

Increasing certainty in projected local extreme precipitation change

Corresponding Author: Dr Chao Li

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This study aims to constrain future changes in extreme precipitation under high-emission scenario by utilizing global warming trends from the historical period as constraint. The authors suggest constrained projections, which indicate a weaker intensification compared to the original climate model simulations, along with reduced uncertainties in future projections. While acknowledging the value of providing quantitative information on these aspects, I have identified several significant concerns. Moreover, a large portion of the results presented in this study appeared to be duplicated from a previous study, including one authored by the same researchers (Li et al 2024; https://doi.org/10.1029/2023GL105605). Honestly, I cannot confidently assert that this manuscript is currently suitable for publication in Nature Communications, at least in its current state. Therefore, I recommend rejecting this study.

Major comments
- Authors presented the impact of constraining analysis in Fig. 5. In the results, the changes in projection uncertainties (i.e., variances, Fig. 5D) are quite small in most global land areas, except for East Asia. This suggests that the constraint analysis framework adopted in this study is not improved compared to previous studies (e.g., Paik et al 2023; Li et al 2024).
Paik et al 2023, Emergent constraints on future extreme precipitation intensification: from global to continental scales, Weather and Climate Extremes, https://doi.org/10.1016/j.wace.2023.100613
Li et al 2024, Constraining Projected Changes in Rare Intense Precipitation Events Across Global Land Regions, Geophysical Research Letters, https://doi.org/10.1029/2023GL105605

- More importantly, the authors suggested that their emergent constraint strategy reduces error variance by 25–70% over most areas of the world, as indicated in Fig. 6A. However, the increased effectiveness of the emergent constraint framework, particularly highlighted in Fig. 6A (compared to Fig. 5D), is primarily attributed to the employment of data-aggregation (averaging the diagnosed changes within 3×3 grid cell regions). It is crucial to keep in mind that comparing the constrained and unconstrained projections, with and without employing data-aggregation, respectively, is inherently unfair. This comparison does not accurately reflect the impact of the emergent constraint; rather, it merely presents the results based on regional averages, which naturally reduces projections uncertainties.

- Many of results presented in this study strongly overlap with those showed in Li et al (2024), although there are some differences between them, such as the analysis period and the focus on small-area averages versus grid-scale analysis. In its current state, I find that the uniqueness of this study may be insufficient for publication in Nature Communications, especially considering that studies by Paik et al (2023) and Li et al (2024) have already been published.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Dear authors and editors,
This research paper, NCOMMS-24-05034, by Li et al., suggests more credible extreme precipitation changes by minimizing the role of internal variability before applying the observed temperature-based emergent constraint. I am delighted to encounter this type of paper, which is novel in constraining future extreme precipitation on not a continental (or regional) but a local scale. From this point of view, this study has distinctiveness compared to the previous studies. Although written English is clear and their results sustain the authors' arguments, I documented some comments for authors to address the possible queries debating between assertions from the authors and others. I believe the contents of this paper are a candidate for publication in Nature Communication after a revision process.

Comments:
1. The authors mentioned the "dynamic contribution," but I am slightly worried about the definition of the dynamic term by residual (changes in precipitation minus the thermodynamic role). As arithmetical interaction (multiplication of two variants) exists apart from the dynamic and thermodynamic contribution (e.g., Huang et al., 2023; Paik et al., 2023), if authors still want to hold "dynamic" as the residual, at least authors should provide the supplementary figure proving negligible scale in "interaction" even in a severe warming case of SSP5-8.5. Otherwise, direct computation of dynamic terms looks more appropriate, as in the papers mentioned earlier.

2. Highlight the most important findings of this study
As the constraints on the local scale are the uniqueness of this study, I recommend that the authors emphasize the reduction effects on the local scale by revising Figure 6. According to Figure 6A, some grids where the uncertainty is reduced effectively match the location of megacities (e.g., Beijing, Shanghai, Tokyo, Paris, and London). These cities can be great candidates to accentuate the impact of this study.

3. Regarding the credible projection suggested in this study, the future projections based on SSP5-8.5 are different from the current climate mitigation policies after the Paris Agreement. This high-emission scenario has merits to warn people of the importance of mitigation plans based on the hypothetical experiments. The fact is that countries strive to alleviate rising temperatures by no more than 2.0 degrees, vastly different from SSP5-8.5 details (e.g., the abrupt decreases in CO2 emissions). A general expectation is that the constraining future uncertainty in lower emission scenarios (like SSP2-4.5) would be less effective than in SSP5-8.5, as Paik et al. (2023) reported.

Moreover, even at the same global warming levels, precipitation increases more in RCP2.6 than in RCP8.5 (Mitchel et al., 2016). Some previous articles introduced the relative importance of forcing conditions under different scenarios, illustrating the role of anthropogenic aerosols as a critical component to constraining future projections in precipitation under low-emission scenarios (e.g., Lee et al., 2023; Wang et al., 2023) which is apart from the GHGs dominant role in SSP5-8.5.

To highlight the authors' findings and as a pathfinder for future studies hereafter, additional discussion on constraining future projection uncertainties for low-emission scenarios is worth considering.

References:
Huang, Z., Tan, X., Gan, T. Y., Liu, B., & Chen, X. (2023). Thermodynamically enhanced precipitation extremes are due to the counterbalancing influences of anthropogenic greenhouse gases and aerosols. Nature Water, 1(7), 614-625.
Paik, S., An, S. I., Min, S. K., King, A. D., & Kim, S. K. (2023). Emergent constraints on future extreme precipitation intensification: from global to continental scales. Weather and Climate Extremes, 42, 100613.
Mitchell, D., James, R., Forster, P. M., Betts, R. A., Shiogama, H., & Allen, M. (2016). Realizing the impacts of a 1.5°C warmer world. Nature Climate Change, 6(8), 735-737.
Lee, D., Sparrow, S. N., Min, S. K., Yeh, S. W., & Allen, M. R. (2023). Physically based equation representing the forcing-driven precipitation in climate models. Environmental Research Letters, 18(9), 094063.
Wang, P., Yang, Y., Xue, D., Ren, L., Tang, J., Leung, L. R., & Liao, H. (2023). Aerosols overtake greenhouse gases, causing a warmer climate and more weather extremes toward carbon neutrality. Nature Communications, 14(1), 7257.

(Remarks on code availability)
The authors declared that the codes would be provided after publication.

Reviewer #3

(Remarks to the Author)
Dear authors,
Please find my comments in the attached PDF.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This study offers constrained projections of future extreme precipitation at local scales, by using CMIP6 model simulations. This analysis was conducted based on emergent constraints and data aggregation. I found that the authors considerably effort in addressing my previous comments. However, I still have major concerns regarding the results presented in this study, as I cannot get clear answers for my concerns. Therefore, I hope to suggest major revision again.

Major Comments
1. The authors employed a data aggregation strategy to present results in Fig. 6, demonstrating the effectiveness of this analysis framework in reducing uncertainties in future extreme precipitation projections. However, as I previously commented, comparing spatial aggregated to each individual grid's projections to illustrate variance reduction is both scientifically unfair and inappropriate. This approach does not align with rigorous scientific standards. It appears to data handling process to artificially present lower uncertainties for a specific purpose. Therefore, I cannot accept these results as scientific conclusions, as they exaggerated the effects of adaptative emergent constraint in the presented results (e.g., the reported 20-70% reduction in future projection uncertainty).

2. Furthermore, the role of aggregating ensemble members in reducing internal variability uncertainty remains unclear. Although not clearly explained, it appears that all future projections are based on multi-model means derived from ensemble averages within each climate model used. In this case, the data aggregation strategy does not, in fact, incorporate the influence of averaging ensemble members to reduce the internal variability uncertainty. Please clarify this point to avoid confusion.

Minor Comments
1. Page 3, Line 16: There is a typo in the reference, which is written as '8–1'.

Reviewer #2

(Remarks to the Author)
I appreciate the authors' efforts to improve the quality of this paper by adding scientifically clear and significant evidence and extensive discussions. This manuscript version will attract more general readers of Nature Communications.

I am pleased with the majority of the authors' responses. However, there are some parts of their response (specifically, 3-2) for 'my previous Comment 3' that I find difficult to accept.

The authors mentioned, "Nevertheless, there are relatively higher aerosol concentrations in the low SSP1-2.6 ..."
However, the aerosol concentration (~emission, considering their lifetime scale) in SSP1-2.6 is much lower than in SSP2-4.5 or SSP5-8.5 (e.g., Figure 1 of Turnock et al., 2020).
Relative to current aerosol concentrations (almost peak for the whole period), their future reduction in scenarios will work as positive radiative forcing (e.g., Figure 3 of Smtih et al., 2021), contributing to increased global warming levels.

Please note that the aerosol reduction is much more severe in SSP1-2.6 than in SSP5-8.5, and it is equivalent that the aerosol contribution to rising temperature is higher for SSP1-2.6 > SSP5-8.5, as shown in Figs. 6.22 & 6.24 of Szopa et al., 2021)

**The net global warming is SSP5-8.5 >> SSP1-2.6, mainly explained by the radiative forcing differences in their CO2 forcings (e.g., Figure 6 in Meinhausen et al., 2020).

Therefore, the description in lines 185-186 of the revised manuscript also looks strange because the net global/local warming differences between SSP1-2.6 and others are not due to aerosols but GHGs (Figure 6.24 in Szopa et al., 2021 & Figure 6 in Meinhausen et al., 2020).
Please note that aerosol concentration gaps between SSP1-2.6 and other scenarios compensate for the differences in global warming caused by the disparity in GHGs.

The authors must check their understanding of aerosol forcing differences in CMIP6 scenarios and revise the manuscript and Fig. S5 caption accordingly.

References:
Turnock, S. T., Allen, R. J., Andrews, M., Bauer, S. E., Deushi, M., Emmons, L., ... & Zhang, J. (2020). Historical and future changes in air pollutants from CMIP6 models. Atmospheric Chemistry and Physics, 20(23), 14547-14579.
Smith, C. J., Harris, G. R., Palmer, M. D., Bellouin, N., Collins, W., Myhre, G., ... & Forster, P. M. (2021). Energy budget constraints on the time history of aerosol forcing and climate sensitivity. Journal of Geophysical Research: Atmospheres, 126(13), e2020JD033622.
Meinshausen, M., Nicholls, Z. R., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., ... & Wang, R. H. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. Geoscientific Model Development, 13(8), 3571-3605.
Szopa, S., V. Naik, B. Adhikary, P. Artaxo, T. Berntsen, W.D. Collins, S. Fuzzi, L. Gallardo, A. Kiendler-Scharr, Z. Klimont, H. Liao, N. Unger, and P. Zanis, 2021: Short-Lived Climate Forcers. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis,

M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 817–922, doi:10.1017/9781009157896.008.

Reviewer #3

(Remarks to the Author)
Dear authors,
Thank you for carefully addressing the comments from my review.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you to the authors for their thorough discussion and revisions. However, I still have two comments/suggestions that I would like to emphasize. I hope the authors seriously think and consider, especially my second comment during the revision process.

1. Figure 5d shows the effect of temperature constraints on future projection uncertainty. In Fig. 6a, it only adds the effect of spatial aggregation on future projection uncertainty. Based on my understanding, both results are based on the multi-model mean of multi-ensemble simulations, and the key difference between Fig. 6a and Fig. 5d is ‘spatial aggregation’, not from ‘data aggregation (i.e., spatial and ensemble aggregation)’. If this understanding is correct, some of the descriptions may not be entirely accurate, which explains the Fig. 6a as the results of reduction of uncertainty achieved through spatial and ensemble data aggregation.

2. Uncertainty reduction is achieved through (i) temperature constraints (which are physically interpretable) and (ii) data aggregation, the latter includes (i) spatial aggregation and (ii) ensemble aggregation. However, I am uncertain ensemble aggregation itself plays a role in reducing uncertainty. More importantly, the spatial aggregation method, aggregating nine grid cells around a grid, is not scientifically sounding. Actually, this is just a data processing than a scientifically valid method for reducing uncertainty. Despite this, the authors have emphasized these results in the abstract, Fig. 6a and Fig. 7. Although it may provide projections with reduced uncertainty, I believe it should be avoided in favor of methods to provide scientifically understandable/meaningful projections. I think the authors may might agree and know this point, as they acknowledge on Page 17 the potential as ‘the cost of biases the estimated extreme precipitation changes'.

Reviewer #2

(Remarks to the Author)
I appreciate the authors’ additional efforts to amend the manuscript. I am pleased to recommend this paper as a publication candidate for Nature Communications.

To respond to the Editor’s request, I address my opinion on the method used in this study, which is the dispute between Reviewer #1 & the authors.
Reviewer #1 raised the question about the validity of results from aggregation. The average results of 3x3 adjacent grids make the percentage changes in each grid smooth, and I understand this smoothing reduces spatial uncertainty related to extreme rainfall (e.g., synoptic scale weather system).
Event rareness for extreme rainfall in limited periods (even several decades) makes sample representativity difficult. If the adjacent grids have similar characteristics to the centered grid, data aggregation is a widely used method, even in an hourly extreme rainfall study (e.g., Park & Min, 2017; Lee et al., 2022), to prepare proper numbers of samples to apply statistical analysis.
Climatologically and geologically, I expect adjacent 3x3 are likely to have a similar distribution in extreme rainfall characteristics in historical periods, which can be the climatological prerequisite for the data aggregation area threshold.
Therefore, to support the validity of data aggregation in 3x3 grids, authors may display the statistical test results (e.g., KS test) comparing the annual maximum rainfall distribution in 30 years from samples of 3x3 (in total 270 samples) and samples of the centered grid point (in total, 30 samples).
Authors can count how many models have significant differences in any grid point. Then, authors can suggest specific conditions for applying 3x3 aggregation (e.g., grids point where less than 10% GCMs have significant differences in KS-test results).
This information can support data aggregation 3x3, which is not just data processing but a statistical method based on the presence of climatological characteristics.

Reference:
Park, I. H., & Min, S. K. (2017). Role of convective precipitation in the relationship between subdaily extreme precipitation and temperature. Journal of Climate, 30(23), 9527-9537.
Lee, D., Min, S. K., Park, I. H., Ahn, J. B., Cha, D. H., Chang, E. C., & Byun, Y. H. (2022). Enhanced role of convection in future hourly rainfall extremes over South Korea. Geophysical Research Letters, 49(22), e2022GL099727.

Reviewer #1 (Remarks to the Author):

This study aims to constrain future changes in extreme precipitation under high-emission scenario by utilizing global warming trends from the historical period as constraint. The authors suggest constrained projections, which indicate a weaker intensification compared to the original climate model simulations, along with reduced uncertainties in future projections. While acknowledging the value of providing quantitative information on these aspects, I have identified several significant concerns. Moreover, a large portion of the results presented in this study appeared to be duplicated from a previous study, including one authored by the same researchers (Li et al 2024; https://doi.org/10.1029/2023GL105605).

Honestly, I cannot confidently assert that this manuscript is currently suitable for publication in Nature Communications, at least in its current state. Therefore, I recommend rejecting this study.

Response:

Thank you for your review of our manuscript. We understand your concerns regarding potential overlap with our previous work (Li et al., 2024), but believe that we clearly outlined the distinction between the present study and previous studies in the Introduction section of our manuscript. Li et al. (2024) and other studies, such as Paik et al. (2023; https://doi.org/10/1016/j.wace.2023.100613) and Dai et al. (2014; https://doi.org/10.1073/pnas.2312400121), focus exclusively on the projections of extreme precipitation at *regional to global scales*, considering areas the size of IPCC regions or larger. As a result, these studies are of only limited value for impact assessment and adaptation planning at local levels. In contrast, our current study provides observationally constrained projections of *local* (1°x1° grid box scale) extreme precipitation, which we believe to be a substantial advance. To the best of our knowledge, there are currently no studies other than our current study that offer observational constraints for projections of local extreme precipitation.

We identified two key challenges that hinder the effectiveness of a temperature-based constraint at the local scale: the substantial influence of unforced internal climate variability and the substantial modulating effect of changes in atmospheric circulations on the response of local extreme precipitation to increases in atmospheric precipitable water. We therefore developed an adaptative local emergent constraint strategy that filters out as much unforced internal climate variability as possible by pooling simulated extreme precipitation from neighboring grid cells and across ensemble simulations of individual climate models, and takes advantage of the robust, strong, and widespread temperature constraints on the thermodynamic component of change in extreme precipitation.

Overall, our study makes two major contributions beyond existing ones. First, we reduced the error variance of projected end-of-century changes in *local* annual extremes of daily precipitation by 20-70% over most areas of the world under various emissions scenarios. This implies that future adaptation investments to limit the risks from extreme precipitation and related hazards can be more effectively planned at finer local levels. Second, our strategy of pinpointing effective observational constraints by decomposing the key components of climate change governed by distinct processes offers a novel pathway for developing emergent constraints on projections of climate variables with low signal-to-noise ratios such as local extreme precipitation. We have emphasized these points further in the revised manuscript.

Major comments:

Comment 1:

Authors presented the impact of constraining analysis in Fig. 5. In the results, the changes in projection uncertainties (i.e., variances, Fig. 5D) are quite small in most global land areas, except for East Asia. This suggests that the constraint analysis framework adopted in this study is not improved compared to previous studies (e.g., Paik et al 2023; Li et al 2024).

Paik et al 2023, Emergent constraints on future extreme precipitation intensification: from global to continental scales, Weather and Climate Extremes, https://doi.org/10.1016/j.wace.2023.100613

Li et al 2024, Constraining Projected Changes in Rare Intense Precipitation Events Across Global Land Regions, Geophysical Research Letters, https://doi.org/10.1029/2023GL105605

Response:

The temperature constraint, as implemented in the referenced studies on extreme precipitation changes at regional and global scales, *cannot be directly applied to local scales.* This point was extensively discussed in our manuscript with strong quantitative evidence. For example, we estimated that the temperature constraint proposed in the referenced studies is valid for annual extremes of daily precipitation over only 10%-35% of the global surface under the considered emissions scenarios (i.e., SSP1-2.6, 2-4.5 and 5-8.5) at a 5% significance level. In contrast, our adaptative constraint is valid across 70%-90% of the global surface at the same significance level. During revision, we further optimized the constraint configuration through a model-based cross-validation analysis (see Figure S15 and the relevant discussions in the Methods section). As a result, our constraint strategy is applicable over 90% of the global surface for the considered emissions scenarios, which covers almost the entire global population based on 2020 statistics.

We believe that expanding the applicability of the simple temperature constraint to globally widespread grid cells with enhanced effectiveness represents a substantial advancement over the referenced studies. Moreover, it offers a distinct methodological framework for establishing emergent constraints on complex climate changes (e.g., low signal-to-noise variables). The adaptative constraint introduced in this study reduces the projection error variance in the forced local end-of-century extreme precipitation changes under the SSP5-8.5 emissions scenario by over 10% across most of the constrained area, with a global median reduction of 18%, even without accounting for reductions in intern variability uncertainty. This is an encouraging improvement for projections of local extreme precipitation, especially considering that local extreme precipitation is notoriously difficult to constrain.

To better highlight the value of these uncertainty reductions, we have included a new figure that illustrates the effect of data aggregation in reducing internal climate variability uncertainty, and the impact of the temperature constraint in reducing forced response uncertainty in extreme precipitation changes for selected grid cells where major cities and economic hubs in North America, Europe, and Asia are located. Please see Figure 7 in the revised manuscript. To understand this figure, it is important to note that the extent to which the temperature constraint can reduce the inter-model spread of the forced extreme precipitation changes in a particular location depends on the contribution of thermodynamic uncertainty to total uncertainty, as shown in Figure 2 of the manuscript. In regions where the dynamic uncertainties are particularly large, the uncertainty reduction achieved by the temperature constraint can be small. We acknowledged this as a limitation of the study, and highlighted the need to identify effective emergent constraints for atmospheric circulation changes relevant to extreme precipitation as a key direction of future research. Please refer to the second last paragraph in the Discussion section of the revised manuscript.

Comment 2:

More importantly, the authors suggested that their emergent constraint strategy reduces error variance by 25–70% over most areas of the world, as indicated in Fig. 6A. However, the increased effectiveness of the emergent constraint framework, particularly highlighted in Fig. 6A (compared to Fig. 5D), is primarily attributed to the employment of data-aggregation (averaging the diagnosed changes within 3×3 grid cell regions). It is crucial to keep in mind that comparing the constrained and unconstrained projections, with and without employing data-aggregation, respectively, is inherently unfair. This comparison does not accurately reflect the impact of the emergent constraint; rather, it merely presents the results based on regional averages, which naturally reduces projections uncertainties.

Response:

Our data aggregation approach pools data across regions containing nine 1°x1° grid boxes that are very small relative to regions used in previous studies, and uses an approach similar to that used frequently by engineers when estimating local return levels from limited station data (see Hoskins and Wallis, 1997, which has received 4262 citations according to Google Scholar as of 24 Sept 2024, and references cited in our manuscript). Restricting the pooling area to only a few grid boxes and centering the pooling area on the location of interest (note that we use moving windows centered on each grid box rather than pooling into discrete non-overlapping areas) ensures that the derived extreme precipitation estimates are as representative of the location of interest as possible.

We aggerate data across small moving windows because internal climate variability can substantially amplify or dampen forced model responses in extreme precipitation, making it challenging to establish robust, physically meaningful emergent constraints on model responses. Aggregating simulations of neighboring grid cells and across ensemble simulations helps reduce this influence at individual grid cells. The internal variability uncertainty reduction that is achieved in turn increases both the global area with an effective temperature constraint on the properties of extreme precipitation, and the overall strength of the constraint.

The uncertainty reduction achieved by our adaptative constraint strategy stems from *two sources*: internal variability uncertainty reduction by data aggregation and forced model response uncertainty reduction via the temperature constraint. These uncertainty reductions were presented separately in different figures, each with an appropriate baseline reference for reduction calculation, as being explicit in relevant discussions in the manuscript. For example, for the projected end-of-century changes in extreme precipitation under the SSP5-8.5 emissions scenario, we presented internal variability uncertainty reduction in Figure S1, model response uncertainty reduction in Figure 5D, and the combined reduction in Figure 6. In the newly added Figure 7, we also illustrate these effects separately. Both data aggregation and the temperature constraint make important contributions to the ultimate uncertainty reduction that is achieved in locations where an effective temperature constraint is found. In addition, the data aggregation approach, which is similar to approaches often used by engineering practitioners, helps to reduce projection uncertainty even in locations where a temperature constraint is not found. We believe our presentation is clear and not misleading.

Comment 3:

Many of results presented in this study strongly overlap with those showed in Li et al (2024), although there are some differences between them, such as the analysis period and the focus on small-area averages versus grid-scale analysis. In its current state, I find that the uniqueness of this study may be insufficient for publication in Nature Communications, especially considering that studies by Paik et al (2023) and Li et al (2024) have already been published.

Response:

This comment repeats concerns raised by the reviewer in his overarching summary and in Major Comments 1 and 2. We therefore reiterate that we disagree with the reviewer's characterization of the degree of overlap with previous studies.

As discussed above, we use data pooling across small moving windows centered on individual grid boxes together with an optimized temperature constraint where feasible to provide constrained local projections of changes in extreme precipitation. Previous studies, including our own, have only been able to provide constraints for median or mean changes in extremes across regions the size of IPCC regions (containing 154-1874 1°x1° grid boxes for different IPCC land regions; Li et al, 2024) or substantially larger regions (Paik et al, 2023),

Moreover, our study clearly demonstrates that transitioning from larger regional scales to more locally relevant scales is not a merely matter of applying methods used at large scales to local scales. In fact, the referenced studies illustrate this point well. The temperature-based constraint proposed in these studies is applicable to only 10%-35% of the global surface under the considered emissions scenarios. In contrast, our adaptative emergent constraint is applicable over 90% of the global surface, covering nearly the entire global population based on 2020 statistics. This advance is achieved by reducing unforced internal climate variability by spatial and ensemble data aggregation, a critical analysis of the thermodynamic and dynamic contributions to projected local extreme precipitation changes, and an optimized constraint configuration. Identifying effective constraints by decomposing the key components of climate change governed by distinct processes offers a novel pathway for developing emergent constraints on climate change projections not just for local extreme precipitation, but also for other climate variables with low signal-to-noise ratios. This approach may stimulate advances in the science of emergent constraints on long-term climate change by offering a distinct methodological framework.

Reviewer #2 (Remarks to the Author):

Dear authors and editors,

This research paper, NCOMMS-24-05034, by Li et al., suggests more credible extreme precipitation changes by minimizing the role of internal variability before applying the observed temperature-based emergent constraint. I am delighted to encounter this type of paper, which is novel in constraining future extreme precipitation on not a continental (or regional) but a local scale. From this point of view, this study has distinctiveness compared to the previous studies. Although written English is clear and their results sustain the authors' arguments, I documented some comments for authors to address the possible queries debating between assertions from the authors and others. I believe the contents of this paper are a candidate for publication in Nature Communication after a revision process.

Response:

We sincerely appreciate your positive assessment of our manuscript and are particularly encouraged by the recognition of the distinctiveness of our work in constraining local extreme precipitation. We provide our detailed responses to your specific comments and suggestions below.

Comment 1:

The authors mentioned the "dynamic contribution," but I am slightly worried about the definition of the dynamic term by residual (changes in precipitation minus the thermodynamic role). As arithmetical interaction (multiplication of two variants) exists apart from the dynamic and thermodynamic contribution (e.g., Huang et al., 2023; Paik et al., 2023), if authors still want to hold "dynamic" as the residual, at least authors should provide the supplementary figure proving negligible scale in "interaction" even in a severe warming case of SSP5-8.5. Otherwise, direct computation of dynamic terms looks more appropriate, as in the papers mentioned earlier.

Response:

Thank you for your feedback, and we appreciate your concern regarding the dynamic component of the projected extreme precipitation changes.

Our decomposition is chosen to make the emergent constraint procedure powerful and manageable. We separated the thermodynamic contribution from increases in atmospheric moisture from other contributions that we collectively referred to as "dynamic" contribution. We appreciate your reminder that the latter could include both direct dynamic influences and contributions from thermodynamic-dynamic interactions, and have now made this clear in the text. It should be noted, however that we only use the thermodynamic component in our adaptive emergent constraint procedure, making further decomposition of the residual term unnecessary.

To prevent potential misunderstanding, we have revised the relevant descriptions in the manuscript. For example, when the decomposition is first introduced in the main text, we have revised the explanation as follows:

"*Based on a physical scaling diagnostic for extreme precipitation intensity* [20]*, the projected extreme precipitation changes can be decomposed into a thermodynamic component due to increases in atmospheric precipitable water and a residual component dominated by changes in atmospheric vertical velocity (Methods), which is referred to as the dynamic component throughout the manuscript.*"

We have performed some additional analysis to confirm that we have not missed something by relegating the contribution of dynamic-thermodynamic interactions to the projected extreme precipitation changes to the residual component. We found that the interaction component is small in magnitude and closely resembles the spatial pattern of dynamic changes, as illustrated in Figure R1, which shows the diagnosed thermodynamic (A), dynamic (B), and thermodynamic-dynamic interaction (C) components of the projected changes in annual maximum daily precipitation during

2071-2100 relative to 1985-2014 under the SSP5-8.5 scenario. This supports combining the interaction component with the dynamic component to define the residual term.

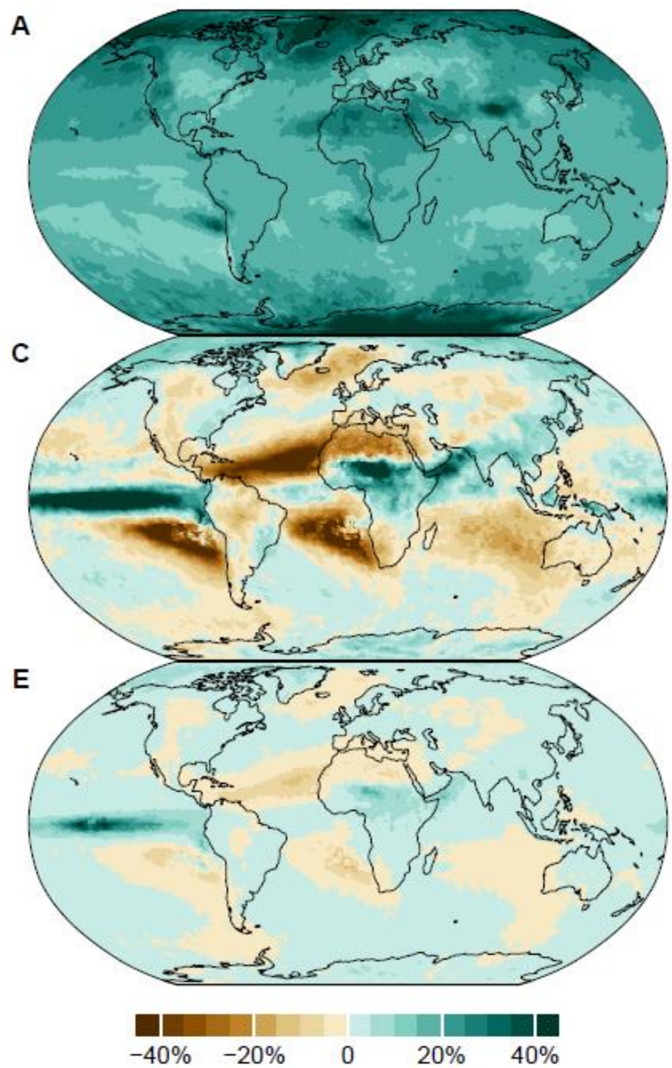

Figure R1. The diagnosed thermodynamic (A), dynamic (B) and thermodynamic-dynamic interaction (C) components of the projected multi-model median changes in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 scenario.

Comment 2: Highlight the most important findings of this study

As the constraints on the local scale are the uniqueness of this study, I recommend that the authors emphasize the reduction effects on the local scale by revising Figure 6. According to Figure 6A, some grids where the uncertainty is reduced effectively match the location of megacities (e.g., Beijing, Shanghai, Tokyo, Paris, and London). These cities can be great candidates to accentuate the impact of this study.

Response:

Thank you for the valuable suggestion, which helps emphasize the importance of our work in reducing uncertainty in projected local extreme precipitation changes. To avoid overloading Figure 6 with additional information, we have added Figure 7 in the revised manuscript. This new figure illustrates the unconstrained and constrained projections of local extreme precipitation changes in selected grid cells where several of the world's largest cities and economic hubs, such as Beijing, Tokyo, Paris, and London, are located. For ease of reference, this figure is also included below. The results have also been briefly highlighted in the manuscript as follows:

"*By permitting more accurate impact assessment and adaptation planning, they are expected to benefit ~90% of the world population [67] (with uncertainty reduction of >10%; Fig. 6B). Specifically, for East and South Asia, which is the home to over 40% of the world's population experiencing climatological heavy monsoon rainfall, the constrained annual maximum daily precipitation projections by late this century are 15% to 70% less uncertain than unconstrained projections. The pronounced uncertainty reductions in regions containing many of the world's largest cities and economic hubs can provide support for advancing climate resilience in these globally important urban areas (Fig. 7).*"

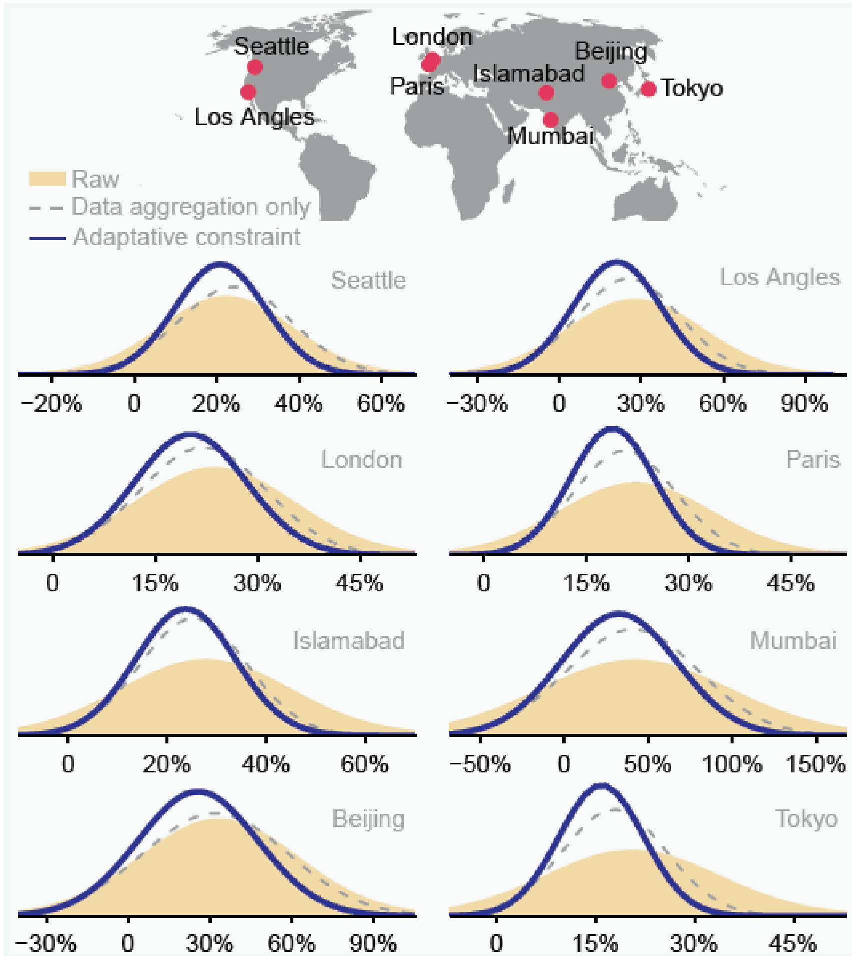

Fig. 7. Constrained projections of future extreme precipitation for selected grid cells. Panels show the probability distributions of the projections of changes in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 forcing scenario for grid cells containing the selected large cities marked on the map, based on unconstrained projections without (yellow) and with data aggregation (gray) and constrained projections using the adaptive constraint method (blue).

Comment 3:

Regarding the credible projection suggested in this study, the future projections based on SSP5-8.5 are different from the current climate mitigation policies after the Paris Agreement. This high-emission scenario has merits to warn people of the importance of mitigation plans based on the hypothetical experiments. The fact is that countries strive to alleviate rising temperatures by no more than 2.0 degrees, vastly different from SSP5-8.5 details (e.g., the abrupt decreases in CO2 emissions). A general expectation is that the constraining future uncertainty in lower emission scenarios (like SSP2-4.5) would be less effective than in SSP5-8.5, as Paik et al. (2023) reported.

Moreover, even at the same global warming levels, precipitation increases more in RCP2.6 than in RCP8.5 (Mitchel et al., 2016). Some previous articles introduced the relative importance of forcing conditions under different scenarios, illustrating the role of anthropogenic aerosols as a critical component to constraining future projections in precipitation under low-emission scenarios (e.g., Lee et al., 2023; Wang et al., 2023) which is apart from the GHGs dominant role in SSP5-8.5.

To highlight the authors' findings and as a pathfinder for future studies hereafter, additional discussion on constraining future projection uncertainties for low-emission scenarios is worth considering.

References:

Huang, Z., Tan, X., Gan, T. Y., Liu, B., & Chen, X. (2023). Thermodynamically enhanced precipitation extremes are due to the counterbalancing influences of anthropogenic greenhouse gases and aerosols. Nature Water, 1(7), 614-625.

Paik, S., An, S. I., Min, S. K., King, A. D., & Kim, S. K. (2023). Emergent constraints on future extreme precipitation intensification: from global to continental scales. Weather and Climate Extremes, 42, 100613.

Mitchell, D., James, R., Forster, P. M., Betts, R. A., Shiogama, H., & Allen, M. (2016). Realizing the impacts of a 1.5°C warmer world. Nature Climate Change, 6(8), 735-737.

Lee, D., Sparrow, S. N., Min, S. K., Yeh, S. W., & Allen, M. R. (2023). Physically based equation representing the forcing-driven precipitation in climate models. Environmental Research Letters, 18(9), 094063.

Wang, P., Yang, Y., Xue, D., Ren, L., Tang, J., Leung, L. R., & Liao, H. (2023). Aerosols overtake greenhouse gases, causing a warmer climate and more weather extremes toward carbon neutrality. Nature Communications, 14(1), 7257.

Response:

Thank you for this thoughtful feedback. We appreciate your concern regarding the relevance of the high SSP5-8.5 emissions scenario, especially in light of the current climate mitigation policies following the Paris Agreement. We also recognize the need to assess the robustness of the adaptative constraint method with respect to other lower emissions scenarios, given the notable dependence of projected changes in mean precipitation (we guess) on the composition of emissions scenarios. In response to these points, we have made the following revisions.

First, we have highlighted the rational for constraining the projections of local extreme precipitation changes under the high SSP5-8.5 emissions scenario in different places of the manuscript, for example:

"*For instance, to protect infrastructure such as roads, bridges, and drainage systems from increasingly intense downpours, planners need accurate projections of future precipitation extremes* [6]*, often with particular emphasis on the potential severe stress that infrastructure may face as informed by projections under high-end emissions scenarios* [7]*.*"

"*We focus on constraining changes in annual maximum daily precipitation accumulation during 2071-2100 relative to 1985-2014 projected to occur in response to a high emission scenario as represented by the Shared Socioeconomic Pathway 5-8.5 (SSP5-8.5)* [34] *in order to highlight the potential severe stress infrastructure may face in this century. In addition, we also consider projections under the lower SSP1-2.6 and SSP2-4.5 emissions scenarios* [34]*.*"

Second, we have also included and discussed the results for the SSP1-2.6 and SSP2-4.5 emissions scenarios. The results reveal that:

1) The spatial and ensemble data aggregation leads to more substantial reductions in internal climate variability uncertainty in the projected extreme precipitation changes under the lower emissions scenarios. This has been reflected in the revised manuscript, as follows:

"*Aggregating data in this way reduces the inter-model variance of the projected changes in total, thermodynamic, and dynamic grid-cell extreme precipitation under the SSP5-8.5 emissions scenario by about 26%, 15%, and 34% on average over the globe, respectively, with even greater reductions for other lower emissions scenarios (Fig. S1).*"

2) The areas with valid temperature constraints and the strength of those constraints slightly reduce under the lower emissions scenarios. We have discussed this in the revised manuscript as follows:

"*Under lower emissions scenarios, substantially less areas exhibit statistically significant inter-model correlations between the projected extreme precipitation total changes and simulated historical GSAT trends (Fig. S3), largely due to the relatively larger influence of unforced internal climate variability. This further highlights the limited utility of a simple temperature constraint at local scales. In contrast, the thermodynamic contributions continue to show significant correlations with historical GSAT trends across the global surface (Fig. S3), albeit over smaller areas and with reduced strength, particularly under the SSP1-2.6 scenario (Fig. S3 vs. Fig. 3).*"

We have further investigated the mechanisms responsible for the attenuated constraints under the lower emissions scenarios. We analyzed the rate of the diagnosed total and thermodynamic changes in annual maximum daily precipitation with respect to global warming under the emissions scenarios considered and have shown results in a supplementary figure, which is included below for ease of reference. As shown, the scaling rate remains rather homogeneous across these scenarios. This means that the projected total and thermodynamic changes in annual maximum daily precipitation are largely determined by the extent of global warming and do not strongly depend on emissions scenarios, as also reported in previous studies (e.g., Pendergrass et al., 2015, https://doi.org/10.1002/2015GL065854; Sillmann et al., 2019, https://doi.org/10.1038/s41612-019-0079-3; Li et al., 2021, https://doi.org/10.1175/JCLI-D-19-1013.1). We note that these results are not in conflict with previous studies that emphasize the role of anthropogenic aerosols in altering extreme precipitation such as the referenced Huang et al. (2023), Park et al. (2023), and Wang et al. (2023). In these studies, aerosols influence long-term changes in extreme precipitation primarily through modulating the strength of warming.

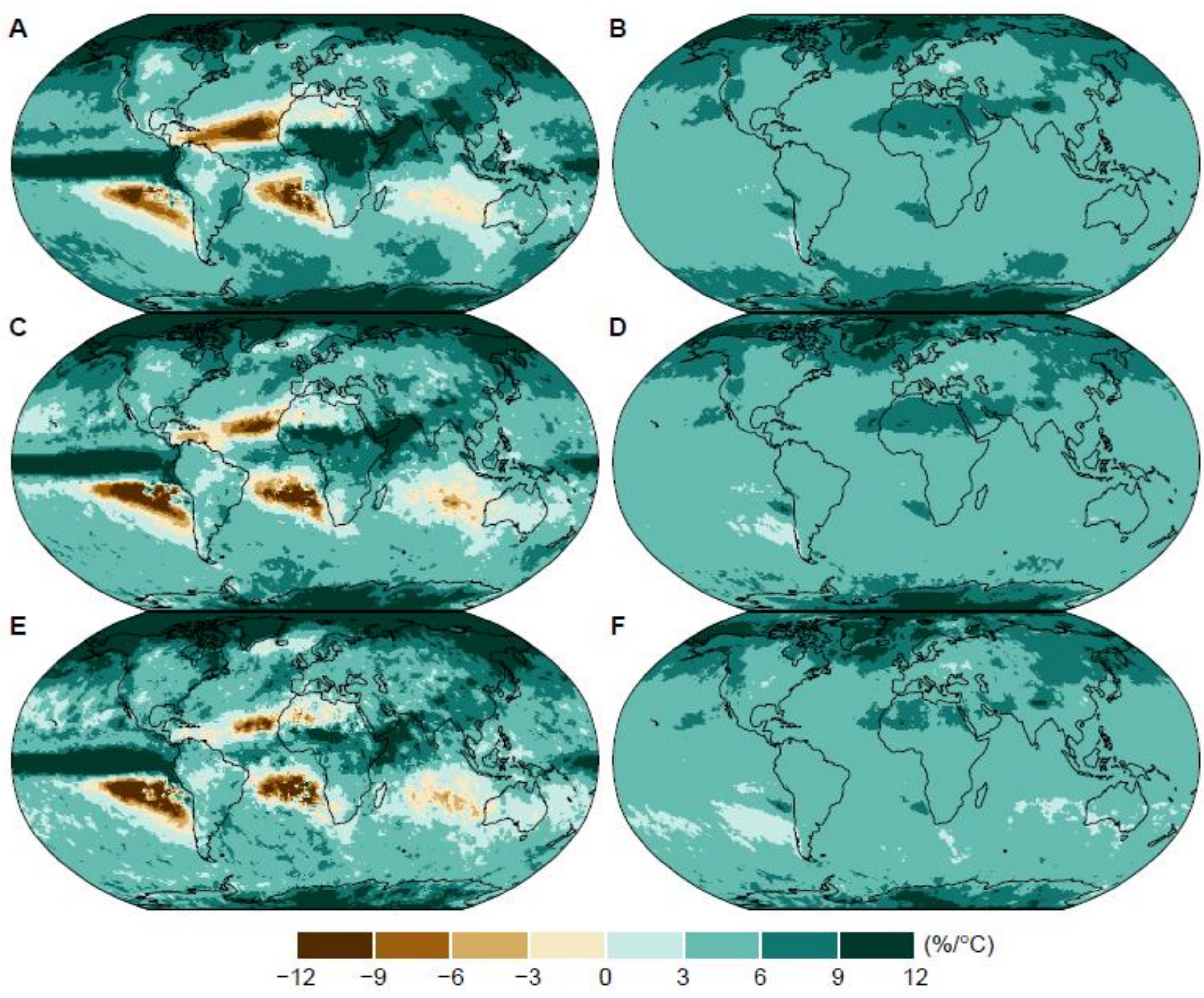

Fig. S4. The projected scaling rates of extreme precipitation changes with global warming. The multi-model median scaling rates of the diagnosed total (**A**, **C**, and **E**) and thermodynamic (**B**, **D**, and **F**) changes in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 with global warming under the SSP5-8.5 (**A**-**B**), SSP2-4.5 (**C**-**D**), and SSP1-2.6 (**E**-**F**) emissions scenarios. The scaling rates do not vary strongly across emissions scenarios, with a few localized exceptions, indicating that the projected annual maximum daily precipitation total and thermodynamic changes depend mainly on the extent of global warming regardless of emissions scenarios.

We believe that the substantially weaker extreme precipitation changes compared to internal climate variability under the lower emissions scenarios greatly obscure the temperature constraint relationships for the projected extreme precipitation. Also, we speculate that the attenuated constraints under the low SSP1-2.6 scenario are partly caused by the weakened temperature constraints on future warming under this scenario. An underlying assumption in using the past warming trend to constraint future climate change is that both the past warming trend and future climate change are driven primarily by greenhouse gas increases. However, under the low SSP1-2.6 scenario, the elevated concentrations of aerosols relative to greenhouse gases can challenge this assumption to some extent, leading to weakened constraints on projected warming and consequently on projected extreme precipitation. The figure below, which shows the emergent constraint relationships for future warming projections under the considered emissions scenarios using the global warming trend during 1971-2020, is in line with this speculation, although the larger influence of internal climate variability on global warming also play an important role. We have also included this figure as the supplementary Figure S5.

We have noted the referenced studies on the influence of emissions composition on mean precipitation. The mechanisms influencing changes in extreme precipitation under different emissions scenarios depend on the severity of extreme precipitation (Sillmann et al., 2019, https://doi.org/10.1038/s41612-019-0079-3). Aerosols can influence mean precipitation and moderate extreme precipitation via mechanisms beyond radiatively forced warming, such as aerosol radiative effects on the energy balance and microphysical effects on cloud droplets (e.g., Stier et al. 2024, https://doi.org/10.1038/s41561-024-01482-6), leading to different mean and moderate extreme precipitation changes even under the same level of global warming. The referenced studies Mitchell et al. (2016) and Lee et al. (2023) are good examples of this. This contrasts with changes in the studied annual maximum daily precipitation.

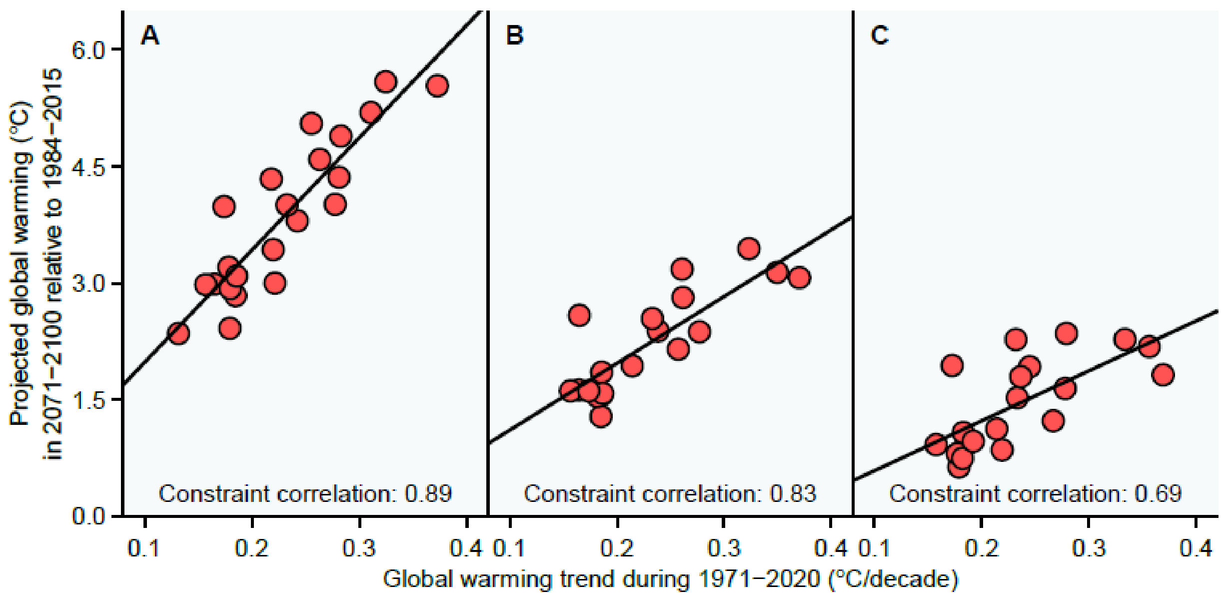

Fig. S5. Scatter plot illustrating the temperature constraint on global warming projections. The scatter points display GSAT changes in 2071-2100 relative to 1985-2014 under the SSP5-8.5 (**A**), SSP2-4.5 (**B**), and SSP1-2.6 (**C**) emissions scenarios versus the 1971–2020 GSAT trends in the studied CMIP6 models (Table S1). In addition to the relatively larger influence of internal climate variability, the remarkably reduced constraint correlation under the SSP1-2.6 scenario is likely due to the relatively elevated concentrations of aerosols compared to greenhouse gases. This challenges the underlying assumption of using the historical warming trend to constrain future warming projections, that is, both the historical warming trend and future climate change are primarily driven by greenhouse gas increases.

The above points have been concisely reflected in the revised manuscript as follows:

"*Under lower emissions scenarios, substantially less areas exhibit statistically significant inter-model correlations between the projected extreme precipitation total changes and simulated historical GSAT trends (Fig. S3), largely due to the relatively larger influence of unforced internal climate variability. … … Unlike mean precipitation [49-51] and moderate extreme precipitation [52-54], the projected total and thermodynamic changes in annual maximum daily precipitation for a given level of global warming do not strongly depend on emissions scenarios [4, 54, 55] (Fig. S4). Nevertheless, the relatively higher aerosol concentrations in the low SSP1-2.6 scenario can weaken a temperature-based constraint on extreme precipitation by reducing its effectiveness in constraining future warming (Fig. S5; Methods), which is directly linked to the thermodynamic component of extreme precipitation changes.*"

3) The temperature constraint also suggests weaker extreme precipitation changes and smaller projection error variance across the constrained area in the lower emissions scenarios, but the total reduction of error variance is comparable to or even greater than that in the SSP5-8.5 scenario.

We have discussed this point in the revised manuscript as follows:

"*The adaptative constraint strategy also suggests globally weaker extreme precipitation changes and smaller error variance in the lower emissions scenarios (Fig. S8). The weakening effect appears more pronounced than in the high SSP5-8.5 scenario, primarily due to the smaller magnitude of unconstrained changes in the lower emissions scenarios. On the other hand, as expected, the weaker temperature constraints (Fig. S3 vs. Fig. 3) lead to milder reductions in projection uncertainty, especially for the SSP1-2.6 scenario (Fig. S8 vs. Fig. 5). Nevertheless, the total reduction of error variance is comparable to or even greater than that observed in the SSP5-8.5 scenario (Fig. S9 vs. Fig. 6). This is consistent with the fact that unforced internal variability dominates* *uncertainty in the projections of extreme precipitation when the external forcing is weak [13, 65, 66], as also discussed previously.*"

Reviewer #3

Paper: "Increasing certainty in projected local extreme precipitation change", Li et al.

The authors propose novel methodology for (1) diagnosing dynamic and thermodynamic changes to precipitation, (2) data aggregation of these changes to boost signal-to-noise, and (3) constraining local projections of extreme precipitation using a temperature-based emergent constraint that reduces model uncertainty. In general, this manuscript is well-written and clearly describes the methods and results. Furthermore, there is clear value with respect to refining local projections of extreme precipitation, since this is the spatial scales at which adaptation takes place.

There are several issues that I would appreciate a response from the authors on.

Response:

We appreciate the positive feedback on manuscript. We are also grateful for your thoughtful comments. We have carefully considered each of them in the revised manuscript. Our detailed point-to-point responses are provided below.

Comment 1: Increased certainty versus "correctness" of projections from uncertain models.

As is often the case with this sort of message, I'm wary of advertising "increased certainty" as a positive. For example, in some cases reduced uncertainty would be a bad thing if the "true" uncertainty were quite large or, more seriously, if the quantity about which the uncertainty refers to is just plain wrong. I appreciate that you validate your methodology rigorously and ensure that, e.g., the nominal coverage is maintained (which I agree confirms that the MME is not overconfident, at least relative to itself). Ultimately, my concern here is that we present the scientific community with "better" projections (i.e., reduced uncertainty) but the projections are wrong. Particularly for local changes to extreme precipitation, models are famously bad at simulating precipitation changes (see, e.g., Polade et al., 2014, 2017; Gershunov et al., 2019, and many of your references) – which I realize is a motivation for the study in the first place. A recent paper even found that within the United States, many GCMs get the sign wrong for simulating changes in summertime precipitation extremes in the historical period (Risser et al., 2024). You acknowledge this point on line 255: "It should also be recognized that while cross-validation against pseudo-observations from global climate models very substantially increases confidence in the constrained projections, they remain subject to the limitations of process representation in global climate models and thus the realism with which the global models simulate extreme precipitation events." I'm not completely sure what to recommend. One thing that I think would be helpful is an expanded discussion on this in the conclusions, making it more explicit that even with "increased certainty" in projections, the projections themselves could be quite wrong.

Response:

Thank you for your feedback and suggestion. We appreciate your concern that the constrained projections are still subject to other sources of bias and uncertainty that are not fully resolved in climate models or adequately addressed by our adaptative constraint method.

We have expanded the discussions for this limitation, as follows:

"*It should, however, be recognized that while our cross-validation against pseudo-observations of future extreme precipitation from global climate models increases confidence in the constrained projections, they remain subject to the limitations of both resolution and process representation in contemporary global climate models [2-5, 68] and thus the realism with which the global models simulate extreme precipitation events and their changes in response to future emissions scenarios. This inherent uncertainty should be carefully considered when using the constrained local extreme precipitation projections for infrastructure adaptation and resilience planning. Moreover, the temperature-based constraint primarily targets and eliminates bias and uncertainty in* *the projected extreme precipitation changes that are directly or indirectly associated with the projected warming or model transient climate response uncertainty.*"

We have also reorganized the Discussion section to better help readers understand the limitations of our work.

Thank you for drawing our attention to the study Risser et al. 2024, which we believe is the one entitled "Anthropogenic aerosols mask increases in US rainfall by greenhouse gases" published in Nature Communications (https://doi.org/10.1038/s41467-024-45504-8). While it is understandable that climate models may simulate extreme precipitation trends with an opposite sign to observations, we would like to note the large role of internal climate variability in the observed evolution of local extreme precipitation over the historical period, as illustrated using initial-condition ensemble simulations in Li et al. (2019; https://doi.org/10.1029/2018EF001001). Therefore, we should not necessarily conclude that the model response to historical forcing is qualitatively inconsistent with observed. In view of this, we decided not to highlight this point in discussion of the limitations of our study.

Comment 2: Treatment of anthropogenic aerosols and the fast precipitation response.

One reason that GCMs have such large uncertainties in simulating local precipitation change is their treatment of anthropogenic aerosols (Hegerl et al., 2015), in both their internal representation of aerosols (e.g., prescribed versus parameterized) as well as the precipitation response to aerosols. As far as I can tell, the only way your method accounts for anthropogenic aerosols is via the slow response to global aerosol forcing as represented by the decreased radiative forcing in the GSAT time series over 1981-2020. This completely ignores the fast precipitation response to aerosols, which has been shown to be significant even in the historical period in both observations (Tao et al., 2012; Risser et al., 2024) and models (see the PDRMIP papers; Samset et al., 2016; Myhre et al., 2017). There is a growing body of papers outlining how and why one must account for aerosols when trying to assess changes to regional precipitation (Hegerl et al., 2015; Sarojini et al., 2016; Persad et al., 2022). The fact that you ignore the fast precipitation response to aerosols feels like a major limitation for this study, and one that is furthermore not acknowledged anywhere in the manuscript.

Response:

Thank you for the critical comment, and for drawing our attention to these studies highlighting the influence of aerosols on mean and extreme precipitation. These studies report that aerosols (and greenhouse gases) can cause changes in mean and extreme precipitation through radiatively forced cooling (and warming) as well as through mechanisms independent of temperature change. The portion of change that is independent of global mean temperature change is often referred to as the "fast response" (e.g., Sillmann et a., 2017, https://doi.org/10.1002/2017GL073229; Bala et al., 2010, https://doi.org/10.1007/s00382-009-0583-y).

We believe that our projected end-of-century changes in extreme precipitation based on transient climate model simulations reflect both the slow and fast responses of extreme precipitation to external forcing. Recall that our adaptive constraint strategy involves decomposing projected forced change in extreme precipitation into thermodynamic and dynamic components, using global mean temperature change to constrain the projected total change, or the thermodynamic component and to combine this with an unconstrained residual term. While some aspects of the fast response may express themselves through the thermodynamic term, our temperature-based constraint primarily targets and eliminates bias and uncertainty in the projected extreme precipitation changes that are directly or indirectly associated with the projected warming, and thus is dominated by the slow response. It follows that the fast component is primarily included in the residual term, if retained, and is evidently entirely present in the constrained projections as no emergent constraint is applied to the residual term.

Since addressing the "fast response" uncertainty is beyond the scope of our study, we have briefly acknowledged it to avoid potential misunderstanding. For example,

- In the Results section: "*Changes in tropospheric vertical motion promoting extreme precipitation evidently occur for a variety of regionally important reasons, such as the eastward shift of ENSO-driven precipitation* [25, 41]*, weakening of the Walker circulation* [42]*, Hadley cell expansion* [43]*, poleward shift of the storm tracks* [44]*, and possibly the rapid response of extreme precipitation-inducing circulations to the direct radiative effects of increases in atmospheric carbon dioxide and aerosols* [45, 46]*.*"
- In the Discussion section: "*Moreover, the temperature-based constraint primarily targets and eliminates bias and uncertainty in the projected extreme precipitation changes that are directly or indirectly associated with the projected warming or model transient climate response uncertainty.*"

For the different responses of mean and extreme precipitation to changes in aerosols and greenhouse gases, please refer to our response to Comment 3 from the second reviewer.

Comment 3: Annual perspective vs. importance of seasonality for extreme precipitation.

I understand why you take an annual perspective on projected changes to local extreme precipitation, but in my view this completely ignores the important seasonal properties of extreme precipitation and, more directly, the ways that the different mechanisms (e.g., large-scale systems, ARs, convection, TCs, etc.) that generate extreme precipitation may be changing in light of external forcing. Importantly, the seasonality piece is critical for climate change adaptation and mitigation, one of the stated reasons for your study in the first place. Could you assess the seasonality in, e.g., your constrained projections in Figure 5?

Response:

Thank you for this thoughtful feedback. We take the annual perspective because, from an applications perspective (e.g., for impacts risk management), the timing of the extreme is at least somewhat less important as it is the largest single value in a year that is most apt to lead to impacts, such as damaged infrastructure. Nevertheless, we did repeat our analysis to consider projected changes in seasonal maximum daily precipitation. The results reveal that the dynamic uncertainty overwhelmingly dominates the uncertainty in future seasonal extreme precipitation projections throughout the world, limiting the effectiveness of a temperature-based constraint at seasonal scales, despite the robust inter-model correlations between the thermodynamic changes in seasonal extreme precipitation and historical GSAT trends. We have highlighted this limitation in the Discussion section, as follows:

"***Further improvement remains desirable in both the tropical and extra-tropical regions and at seasonal scales.*** *A temperature-based emergent constraint is largely ineffective for the forced extreme precipitation changes in the tropics due to the very large dynamic uncertainties. The dynamic uncertainty is also an important component of the total extreme precipitation projection uncertainty in extratropical regions, with magnitudes that rival or exceed uncertainty in the thermodynamic component of extreme precipitation.* ***In addition, the dynamic uncertainty overwhelmingly dominates the uncertainty in seasonal extreme precipitation projections throughout the world (Fig. S10), limiting the effectiveness of a temperature constraint at seasonal scales, despite the robust inter-model correlations between the thermodynamic changes in seasonal extreme precipitation and historical GSAT trends (Fig. S11)****. Thus, further improvement will ultimately require improvement in simulating the dynamical responses of extreme precipitation to external forcing and the identification of effective emergent constraints for atmospheric circulation changes relevant for extreme precipitation.*"

The cited figures for seasonal extreme precipitation are provided below for ease of reference.

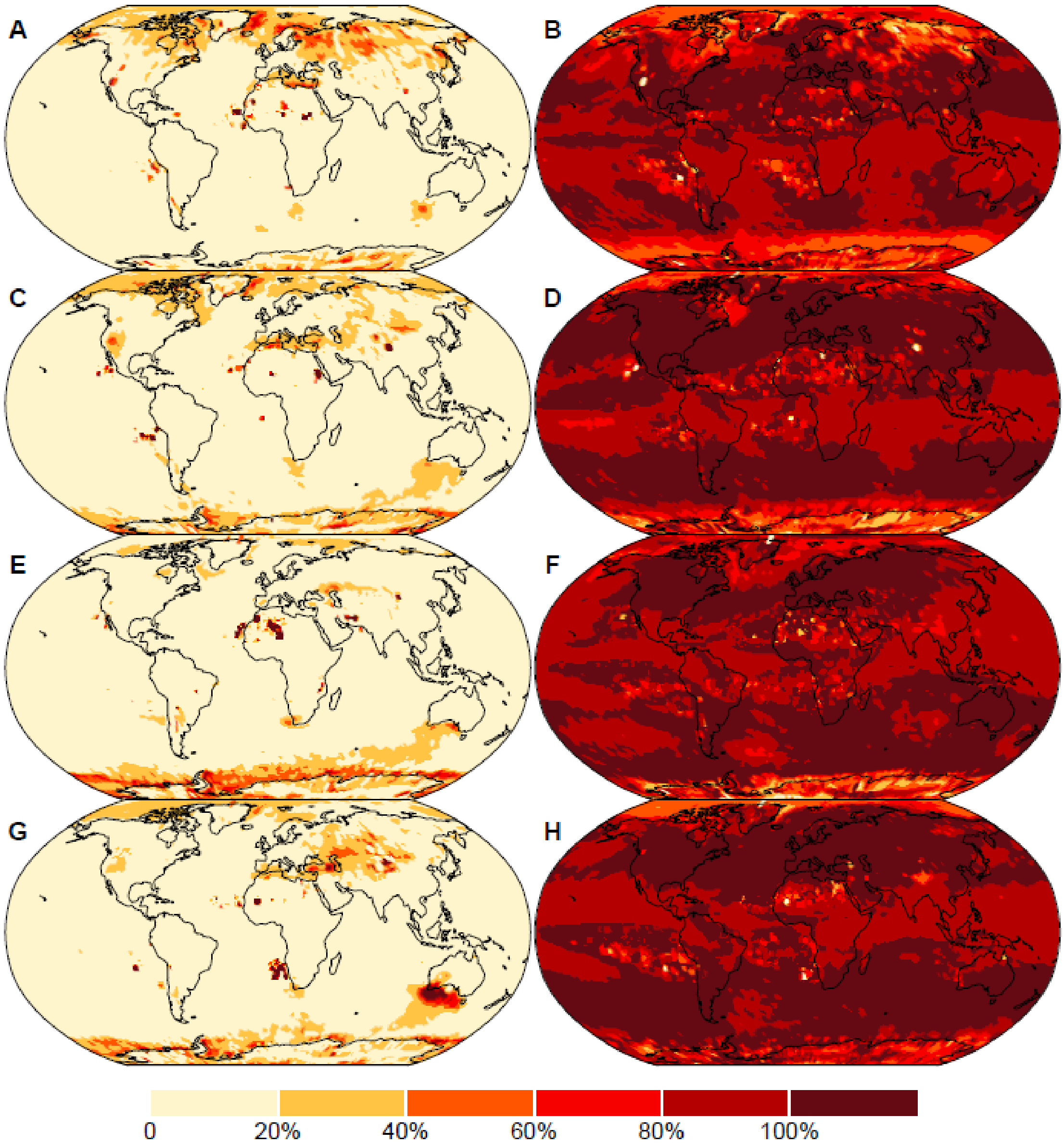

Fig. S10. Contributions of thermodynamic and dynamic uncertainties to total uncertainty at the seasonal scale. The percentage contributions of inter-model thermodynamic (**A**, **C**, **E**, and **G**) and dynamic (**B**, **D**, **F**, and **H**) variances to the total variance of the diagnosed changes in Dec-Jan-Feb (**A-B**), Mar-Apr-May (**C-D**), Jun-Jul-Aug (**E-F**), and Sep-Oct-Nov (**G-H**) maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 emissions scenario. The percentage contributions can be larger than 100% in regions where the thermodynamic and dynamic changes are negatively correlated.

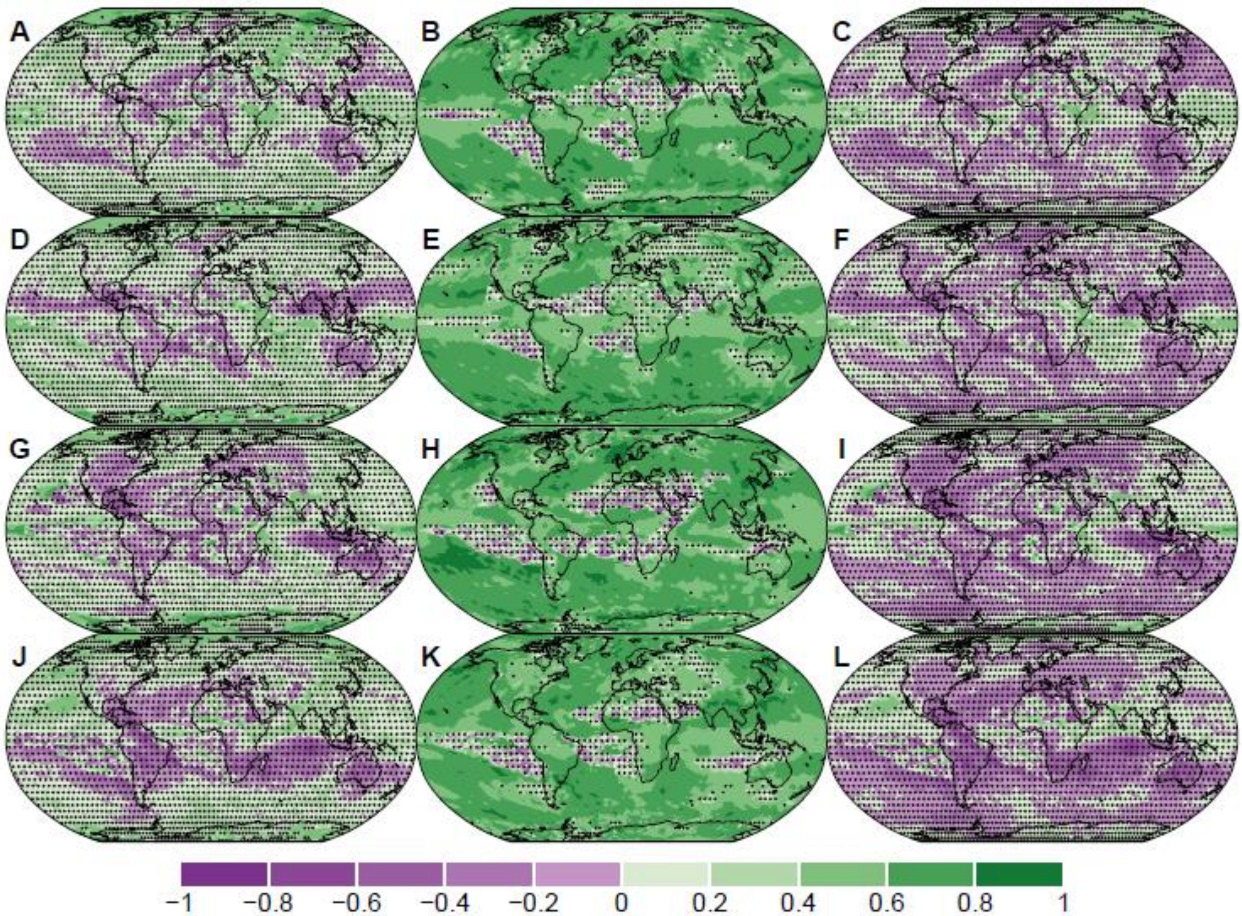

Fig. S11. Correlations between historical warming and future extreme precipitation changes at the seasonal scale. The inter-model correlations between the GSAT trends during 1971-2020 and the diagnosed total (**A**, **D**, **G**, and **J**), thermodynamic (**B**, **E, H**, and **K**), and dynamic (**C**, **F**, **I,** and **L**) changes in Dec-Jan-Feb (**A**-**C**), Mar-Apr-May (**D**-**F**), Jun-Jul-Aug (**G**-**I**), and Sep-Oct-Nov (**J**-**K**) maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 emissions scenario. Stippling marks correlations that are not statistically significant at the 5% level.

Comment 4: Grid-cell testing/stippling

Throughout your manuscript, there are many cases where you conduct a statistical test at each 1◦ ×1◦ GCM grid box: e.g., Figures 3 and 4, and also (I think) in deciding where the adaptive constraint applies in lines 167ff. Conducting a large number of spatially-correlated hypothesis tests as in, e.g., Figure 3 is subject to the usual issues associated with multiplicity in testing (Benjamini and Hochberg, 1995; Wilks, 2016). There are simple ways to address this – see, e.g., Wilks (2016).

Response:

The implementation of the adaptative constraint method requires us to make decisions based on statistical evidence for each individual grid box. This is distinct from the kind of local testing that is performed, for example, to determine whether there is a field significant response to forcing. The latter certainly does require attention to the type of multiple testing issue that is addressed, for example, through the Wilks (2016) false discovery rate procedure or the Livezey and Chen (1983; https://doi.org/10.1175/1520-0493(1983)111%3C0046:SFSAID%3E2.0.CO;2) Monte Carlo procedure. The pertinent question in our study is about local significance rather than field significance, and thus the multiple testing aspect of concern relates to the rate at which incorrect decisions are made, and whether incorrectly made decisions have the potential to seriously impact adaptive constraint performance.

The constraint configuration decision at each individual grid box involves two statistical tests that are applied in sequence, as shown in a newly added schematic diagram for the adaptative constraint method in Fig. S14, which is also included below for ease of reference. The decision procedure is analogous to many applications in climatology that require the construction of decision trees. The first statistical test evaluates whether the inter-model correlation between historical warming and future extreme precipitation total change is significantly positive, while simultaneously the thermodynamic-dynamic covariance is positive. The second statistical test evaluates whether the inter-model correlation between historical warming and future thermodynamic change is significantly positive.

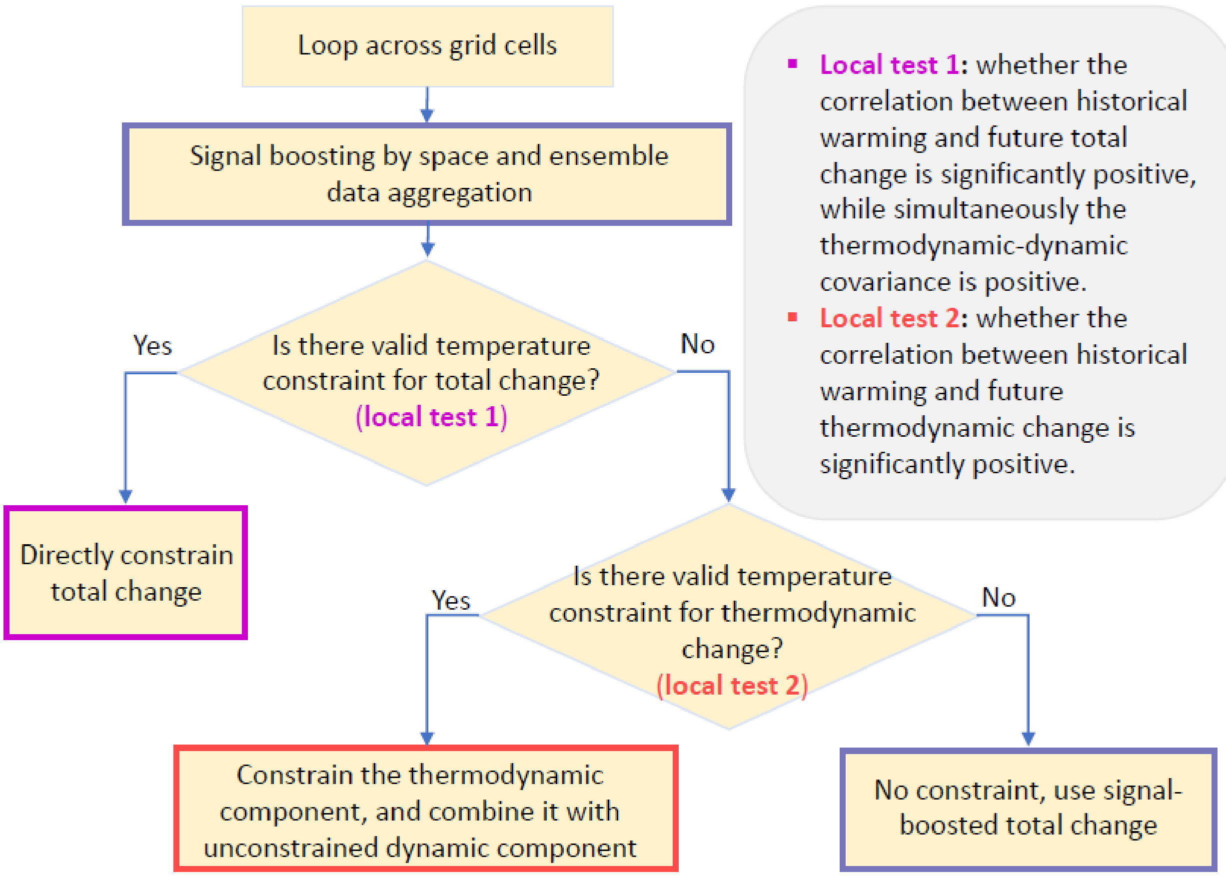

Fig. S14. A schematic diagram illustrating the adaptative constraint method. The method consists of two steps: the first step reduces unforced internal climate variability by aggregating spatial and ensemble simulations; the second step constrains either total or thermodynamic change by a temperature-based constraint, determined by a decision tree that applies two sequential statistical tests. A 10% significance level is used for the statistical tests according to a cross-validation analysis (see Methods).

The reliability of decision making for the constraint configuration depends on both the type I and type II error rates of the two tests. Type I error refers to applying a constraint erroneously when the true constraint correlation is zero. Type II error refers to not applying a constraint when the true constraint correlation is positive, but not detected to be statistically significant. In practice, one can control the type I error rate by altering the given significance level. This will in turn alter the type II error rate such that accepting a higher type I error rate will result in a decrease in the type II error rate. The implication of making a type I error is that a constraint is applied when it would be expected to be ineffective, which is an outcome that would have little impact on the projected change. In contrast, making a type II error implies that an opportunity to effectively constrain the projected change in extreme precipitation is lost. In general, therefore, it should be advantageous to accept a somewhat higher type I error rate, that is, to use a higher significance level.

The overall reliability of the constraint configuration that balances the trade-off between type I and II error rates can be optimized by studying a metric such as the global RMSE in a cross-validation framework and examining how it responds to changes in the significance level (i.e., type I error rate) for the two tests. We have repeated the cross-validation analysis using significance levels ranging from 2.5% to 25%. For simplicity, we applied the same significance level to both local tests. In general, increasing the significance level expands the constrained area and the area where the temperature constraint applies to total change, as shown in the newly added supplementary Figure S15, which is also provided below for ease of reference. In addition, the global mean reduction of RMSE initially increases as the significance level increases and then roughly plateaus (Figure S15C). These results are consistent with expectations. We believe that this robustness of the simple temperature constraint is mainly ensured by the strong thermodynamic foundation of extreme precipitation intensification. Based on these results, we have decided to use a 10% significance level for local testing as a reasonable trade-off that reduces the impact of type II errors while not making type I errors excessively often.

We have discussed these results in the Methods section of the revised manuscript, as follows:

"A schematic flow of the adaptative constraint strategy can be found in Fig. S14. At each grid cell, the constraint configuration involves two sequential statistical tests. The first test evaluates whether the inter-model correlation between the historical warming trend and future extreme precipitation total change is positive at a given significance level, while the thermodynamic and dynamic components interact positively. The second test evaluates whether the inter-model correlation between the historical warming trend and the thermodynamic change is positive at a given significance level. In general, increasing the significance level expands the constrained area and the area where the temperature constraint applies to total change (Figure S15A vs S15B). The reliability of the constraint configuration depends on both the type I and type II error rates of the two tests. The former refers to applying a constraint in circumstances when the constraint would be ineffective, while the latter refers to failing to apply a constraint when it would be effective. Type I errors therefore have little impact, while type II errors lead to unconstrainted local projections that could have been constrained.

We optimized the overall reliability of the adaptative constraint strategy by analyzing the global root mean squared error (RMSE) of constrained extreme precipitation projections with different significance levels in a model-based cross-validation framework, which will be detailed later. Overall, the global mean RMSE reduction increases with larger significance levels and then plateaus (Figure S15C), suggesting that incorrect constraint configuration decisions associated with these varying significance levels have little impact on the performance of the constraint strategy. We believe that this robustness is ensured by the strong thermodynamic foundation of extreme precipitation intensification. Based on these results, we chose a 10% significance level for the involved statistical testing as a reasonable trade-off that reduces the impact of type II errors on projection bias and variance while not making type I errors excessively often."

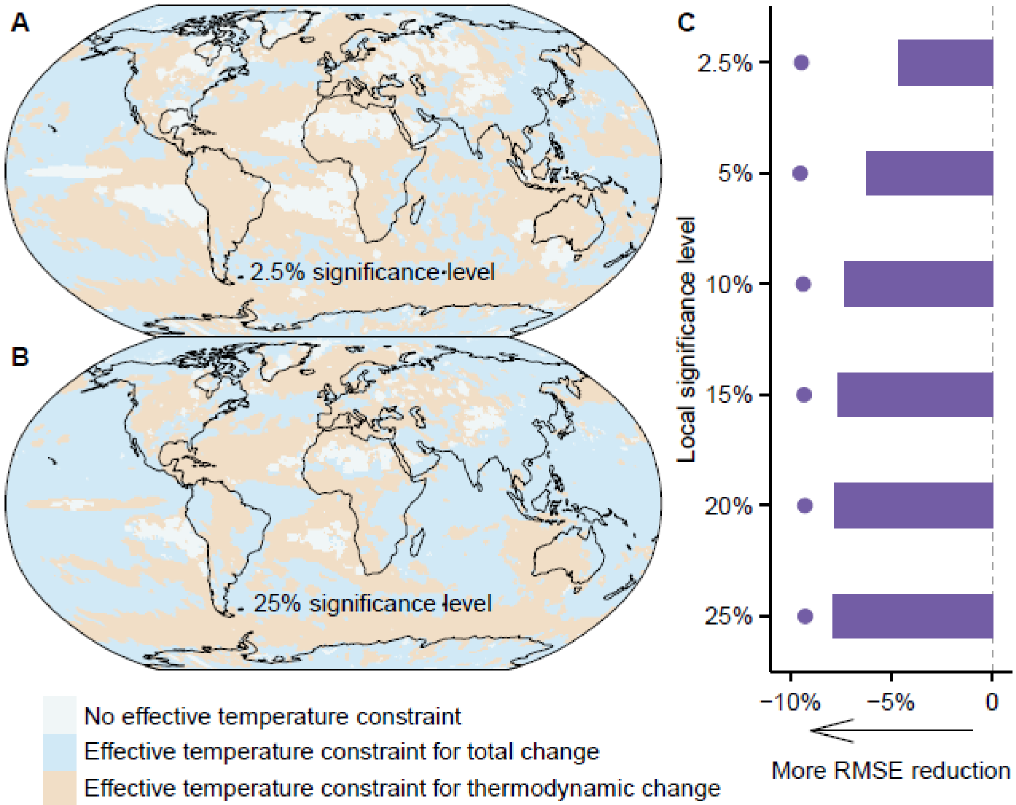

Fig. S15. Optimal configuration analysis for the adaptative constraint method. (**A-B)** The constraint configurations for projected changes in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 scenario obtained using significance levels of 2.5% (**A**) and 25% (**B**) for statistical testing in the adaptative constraint method. (**C**) The global mean (bars) and median (points) percentage differences in root mean squared error between constrained and unconstrained changes in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 scenario, as revealed by model-based cross-validation with significance levels ranging from 2.5% to 25% for constraint configuration determination.

Otherwise, just a few minor concerns:

Minor comment 1:

(Line 194) In what sense is the RMSE a combination of bias and uncertainty? Squared errors are just an assessment of the bias – this is evident in Figure 4, where the “bias reduction” and “RMSE reduction” have nearly identical patterns.

Response:

We have added a note for RMSE as follows:

“*Similarly, we compare the constrained and unconstrained variances and RMSEs (i.e., the square root of the sum of the square of the projection bias and error variance) to ascertain whether the constrained changes have reduced uncertainty intervals and combinations of bias and uncertainty, respectively.*”

Minor comment 2:

Also regarding Figure 4: a small issue with the terminology, although this may be a personal bias. The y-axis in each subpanel is labeled “reduction” – I think “change” might be better? As is, you have a negative reduction, which could imply that the bias, uncertainty, etc. is actually increased.

Response:

Thanks. The suggested change has been made.

Reviewer #1 (Remarks to the Author):

This study offers constrained projections of future extreme precipitation at local scales, by using CMIP6 model simulations. This analysis was conducted based on emergent constraints and data aggregation. I found that the authors considerably efforts in addressing my previous comments. However, I still have major concerns regarding the results presented in this study, as I cannot get clear answers for my concerns. Therefore, I hope to suggest major revision again.

Response:

Thank you for your feedback. We provide our responses to your specific comments below.

Major comments:

Comment 1:

The authors employed a data aggregation strategy to present results in Fig. 6, demonstrating the effectiveness of this analysis framework in reducing uncertainties in future extreme precipitation projections. However, as I previously commented, comparing spatial aggregated to each individual grid's projections to illustrate variance reduction is both scientifically unfair and inappropriate. This approach does no align with rigorous scientific standards. It appears to data handling process to artificially lower uncertainties for a specific purpose. Therefore, I cannot accept these results as scientific conclusions, as they exaggerated the effects of adaptative emergent constraint in the presented results. (e.g., the reported 20-70% reduction in future projection uncertainty).

Response:

This comment repeats concerns raised by the reviewer in his previous review. We believe that the reviewer misinterprets our adaptative constraint method as exclusively referring to the temperature emergent constraint on forced extreme precipitation changes. As we emphasized in our previous responses, the total uncertainty reduction achieved by our constraint strategy arises from two sources: internal variability uncertainty reduction by data aggregation and forced response uncertainty reduction via the temperature constraint. We believe that we did transparently attribute uncertainty reductions to these two sources. Nevertheless, to ensure that their contributions to uncertain reduction are communicated as clearly as possible, we have made several revisions to the paper.

In the caption of Fig. 4, we have specified that:

> "*The validation is based on extreme precipitation changes estimated using spatial and ensemble data aggregation, and thus* ***the bias and uncertainty changes are only due to the application of temperature constraint****.*"

In the caption of Fig 5:

> "*Bias and uncertainty reductions presented in (****c****-****d****) are* ***achieved solely by the temperature emergent constraint and do not include changes due to the application of data aggregation****.*"

In Fig. 6, we have added the corresponding population proportions benefiting from uncertainty reduction achieved by the temperature emergent constraint alone:

> "*(****b****) The proportions of the global population (in the year 2020) residing in grid cells where the total uncertainty reduction is above the levels marked along the horizontal axis.* ***The corresponding population proportions for uncertainty reduction achieved by the temperature emergent constraint alone are shown by red empty bars.***"

In the caption of Fig 7:

> *"… based on unconstrained projections without data aggregation (yellow), unconstrained projections with data aggregation (gray), and constrained*

projections obtained by applying the adaptive temperature emergent constraint to unconstrained projections but with data aggregation (blue)."

Additionally, we have adjusted our presentation where appropriate to avoid potential misinterpretations, as follows:

In the Results section:

"*The projected increases in extreme precipitation are reduced by 13% on average across the constrained area, with reductions of 6% or more in 80% of the area. We also find notable uncertainty reductions in the constrained projections globally, with a global median decrease in projection error variance of 18% and decreases of 5% or more in 80% of the constrained area (Fig. 5D).*"

Note that the context here implies clearly that these changes are achieved by the temperature constraint alone and do not include internal variability uncertainty reduction. Also, in the results section, we now write:

"…. Including the reduction of internal variability uncertainty achieved by spatial and ensemble data aggregation, the global median total reduction of projection error variance increases to 42%, with reductions of 20% or more across 80% of the constrained area (Fig. 6a).

The total uncertainty reduction is calculated by comparing the projection error variance in extreme precipitation changes estimated with data aggregation and then constrained by the temperature emergent constraint strategy to that of unconstrained changes estimated without data aggregation. *Thus, for grid cells where the adaptative temperature emergent constraint is not applicable (e.g., stippled grid cells in Fig. 6a), the total uncertainty reduction comes completely from the reduced internal variability uncertainty obtained by data aggregation."*

In the Discussion section:

> "*Globally, the temperature constraint projects end-of-century changes in annual maximum daily precipitation amount under a high SSP5-8.5 forcing scenario that are 12% smaller than raw, unconstrained projections, with an 18% reduction in projection error variance.* ***Combined with the effect of data aggregation in reducing internal variability uncertainty, this error variance reduction increases further to 42%.***"

In the Abstract:

> "*We present* ***a temperature-based adaptative emergent constraint strategy combined with data aggregation*** *to reduce the error variance of projected end-of-century changes in annual extremes of daily precipitation under a high emissions scenario by >20% over most areas of the world.*"

Given the fact that strong internal climate variability is a critical factor limiting the applicability of a temperature constraint on extreme precipitation changes at local levels, it is necessary, and scientifically justifiable, to reduce projection uncertainty via the application of both data aggregation and an emergent constraint when considering projected changes in extreme precipitation at the local scale. Therefore, it is natural to consider these two types of uncertainty reduction separately, as well as their combined effect, which includes internal variability uncertainty reduction achieved by data aggregation and forced response uncertainty reduction achieved by the temperature emergent constraint.

We obtained a global median uncertainty reduction of 18% by the temperature constraint alone, and a global median total uncertainty reduction of 42%. These represent substantial reductions of the uncertainty in local extreme precipitation projections, even without accounting for the reduction of internal variability uncertainty by data aggregation. Furthermore, we believe that such substantial reductions should be quite encouraging for impacts and adaptation applications.

For internal variability uncertainty extending beyond a decade or so, the primary method to reduce its effects is to obtain a large sample of realizations of internal variability. This can be obtained from large ensemble initial-condition simulations from individual climate models (see Deser et al., 2012, https://doi.org/10.1038/nclimate1562) and/or appropriate spatial pooling (see, for example, Li et al., https://doi.org/10.1029/2018EF001001, and Hoskins and Wallis, 1997, which has received 4262 citations according to Google Scholar as of 24 Sept 2024). Ensemble aggregation effectively reduces internal variability uncertainty across various timescales, while spatial aggregation helps reduce high-frequency internal climate variability uncertainty at local scales. Taken together, they allow us to obtain more accurate estimates of forced local extreme precipitation changes. Undoubtedly, the data aggregation approach is supported by strong scientific and statistical foundations.

Comment 2:

Furthermore, the role of aggregating ensemble members in reducing internal variability remains unclear. Although not clearly explained, it appears that all future projections are based on multi-model means derived from ensemble averages within each climate model used. In this case, the data aggregation strategy does not, in fact, incorporate the influence of averaging ensemble members to reduce the internal variability uncertainty. Please clarify this point to avoid confusion.

Response:

For each climate model, our data aggregation strategy estimates extreme precipitation changes for a given grid cell by pooling data within 3x3 grid cells surrounding the grid cell of interest from all ensemble members of that climate model, as was clearly described in the Methods section, as follows:

> "*In order to reduce internal climate variability in local extreme precipitation, we average the diagnosed changes within 3x3 grid cell regions centered on a given grid cell as well as those from all available initial-condition simulation members of a climate model to obtain diagnosed changes for that grid cell and model. The spatial data aggregation helps reduce the influence of high-frequency internal*

climate variability at small spatial scales [4, 13, 14], while the ensemble aggregation helps reduce the influence of internal climate variability at various time scales, with the extent of the reduction depending on the number of available initial-condition simulations by individual models [35, 69]."

We don't think it is necessary to further separate the role of ensemble pooling from that of spatial pooling.

Minor comments:

Comment 1:

Page 3, Line 16: There is a typo in the reference, which is written as "8-1".

Response:

Thanks. Corrected.

Reviewer #2 (Remarks to the Author):

I appreciate the authors' efforts to improve the quality of this paper by adding scientifically clear and significant evidence and extensive discussions. This manuscript version will attract more general readers of Nature Communications.

I am pleased with the majority of the authors' responses. However, there are some parts of their response (specifically, 3-2) for 'my previous Comment 3' that I find difficult to accept.

The authors mentioned, "Nevertheless, there are relatively higher aerosol concentrations in the low SSP1-2.6 ..." However, the aerosol concentration (~emission, considering their lifetime scale) in SSP1-2.6 is much lower than in SSP2-4.5 or SSP5-8.5 (e.g., Figure 1 of Turnock et al., 2020). Relative to current aerosol concentrations (almost peak for the whole period), their future reduction in scenarios will work as positive radiative forcing (e.g., Figure 3 of Smtih et al., 2021), contributing to increased global warming levels.

Please note that the aerosol reduction is much more severe in SSP1-2.6 than in SSP5-8.5, and it is equivalent that the aerosol contribution to rising temperature is higher for SSP1-2.6 > SSP5-8.5, as shown in Figs. 6.22 & 6.24 of Szopa et al., 2021)

The net global warming is SSP5-8.5 >> SSP1-2.6, mainly explained by the radiative forcing differences in their CO2 forcings (e.g., Figure 6 in Meinhausen et al., 2020).

Therefore, the description in lines 185-186 of the revised manuscript also looks strange because the net global/local warming differences between SSP1-2.6 and others are not due to aerosols but GHGs (Figure 6.24 in Szopa et al., 2021 & Figure 6 in Meinhausen et al., 2020). Please note that aerosol concentration gaps between SSP1-2.6 and other scenarios compensate for the differences in global warming caused by the disparity in GHGs.

The authors must check their understanding of aerosol forcing differences in CMIP6 scenarios and revise the manuscript and Fig. S5 caption accordingly.

References:

Turnock, S. T., Allen, R. J., Andrews, M., Bauer, S. E., Deushi, M., Emmons, L., ... & Zhang, J. (2020). Historical and future changes in air pollutants from CMIP6 models. Atmospheric Chemistry and Physics, 20(23), 14547-14579.

Smith, C. J., Harris, G. R., Palmer, M. D., Bellouin, N., Collins, W., Myhre, G., ... & Forster, P. M. (2021). Energy budget constraints on the time history of aerosol forcing and climate sensitivity. Journal of Geophysical Research: Atmospheres, 126(13), e2020JD033622.

Meinshausen, M., Nicholls, Z. R., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., ... & Wang, R. H. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. Geoscientific Model Development, 13(8), 3571-3605.

Szopa, S., V. Naik, B. Adhikary, P. Artaxo, T. Berntsen, W.D. Collins, S. Fuzzi, L. Gallardo, A. Kiendler-Scharr, Z. Klimont, H. Liao, N. Unger, and P. Zanis, 2021: Short-Lived Climate Forcers. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 817–922, doi:10.1017/9781009157896.008.

Response:

We sincerely appreciate your positive assessment of our revision. We are also grateful for your clarification regarding our misinterpretation of aerosol concentrations in the emissions scenarios. After careful evaluation, we suspect that the reduced constraint correlation for global warming and extreme precipitation under the SSP1-2.6 scenario is evidence that a constraint based on the global mean temperature trend during a historical period of rapid transient response to forcing may be less effective in scenarios that lead to the stabilization of global mean temperature in the latter half of the century. We have made appropriate changes to the caption of Fig S5 and revised the corresponding discussions in the main text, included below for ease of reference.

> "*In addition to the larger influence of internal climate variability, the reduced constraint correlation under the SSP1-2.6 scenario may indicate that a constraint based on the global mean temperature trend during a historical period of rapid transient response to forcing could be less effective in scenarios that lead to the stabilization of global mean temperature during the study period 2071-2100 (Fig. S5).*"

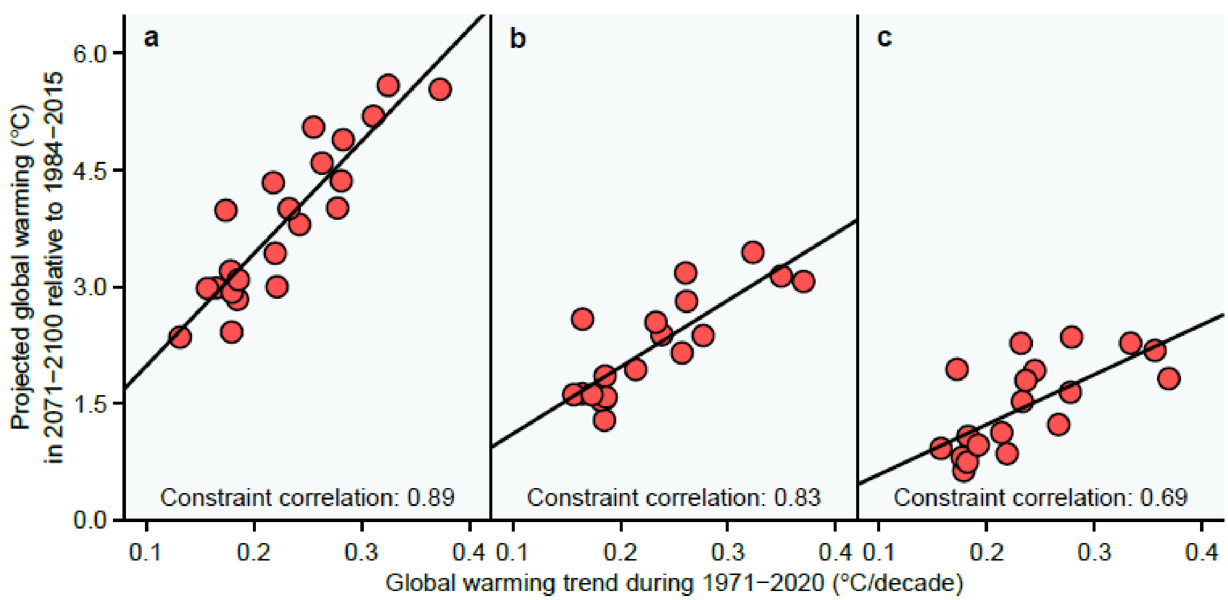

Fig. S5. Scatter plot illustrating the temperature constraint on global warming projections. *The scatter points display GSAT changes in 2071-2100 relative to 1985-2014 under the SSP5-8.5 (**a**), SSP2-4.5 (**b**), and SSP1-2.6 (**c**) emissions scenarios versus the 1971–2020 GSAT trends in the studied CMIP6 models (Table S1).* ***In***

addition to the relatively larger influence of internal climate variability, the reduced constraint correlation under the SSP1-2.6 scenario may be due to the stabilization of global mean temperature in the latter half of the century. This challenges the underlying assumption of using the historical warming trend during a period of transient climate response to constrain future warming projections under stabilization scenarios.

Reviewer #1 (Remarks to the Author):

Thank you to the authors for their thorough discussion and revisions. However, I still have two comments/suggestions that I would like to emphasize. I hope the authors seriously think and consider, especially my second comment during the revision process.

Response:

Thank you for your further feedback.

Comment 1:

Figure 5d shows the effect of temperature constraints on future projection uncertainty. In Fig. 6a, it only adds the effect of spatial aggregation on future projection uncertainty. Based on my understanding, both results are based on the multi-model mean of multi-ensemble simulations, and the key difference between Fig. 6a and Fig. 5d is 'spatial aggregation', not from 'data aggregation (i.e., spatial and ensemble aggregation)'. If this understanding is correct, some of the descriptions may not be entirely accurate, which explains the Fig. 6a as the results of reduction of uncertainty achieved through spatial and ensemble data aggregation.

Response:

In the manuscript, we explicitly defined "data aggregation" as aggregating data from "adjacent grid cells ***and*** all ensemble members from individual climate models". Accordingly, it is natural to interpret "without data aggregation" as using data from individual grid cells and a single ensemble member of a climate model.

Figure 5d shows the percentage reductions in the variance of the constrained changes relative to those of unconstrained changes estimated with data aggregation, and Figure 6a show the percentage variance reductions of the constrained changes relative to those of unconstrained changes estimated without data aggregation. While we believe that our descriptions are clear and accurate, we have revised the related descriptions in Figure 6 to be entirely explicit as follows:

"(a) The percentage variance reductions of the constrained changes in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 forcing scenario compared to those of unconstrained changes estimated without data aggregation ***(i.e., using data from individual grid cells and a single simulation member of a climate model)*** *and thus with high unforced internal climate variability.*"

For the separate effects of spatial and ensemble aggregation on the reduction of internal climate variability uncertainty, please refer to our response to your Comment 2.

Comment 2:

Uncertainty reduction is achieved through (i) temperature constraints (which are physically interpretable) and (ii) data aggregation, the latter includes (i) spatial aggregation and (ii) ensemble aggregation. However, I am uncertain ensemble aggregation itself plays a role in reducing uncertainty. More importantly, the spatial aggregation method, aggregating nine grid cells around a grid, is not scientifically sounding. Actually, this is just a data processing than a scientifically valid method for reducing uncertainty. Despite this, the authors have emphasized these results in the abstract, Fig. 6a and Fig. 7. Although it may provide projections with reduced uncertainty, I believe it should be avoided in favor of methods to provide scientifically understandable/meaningful projections. I think the authors may might agree and know this point, as they acknowledge on Page 17 the potential as 'the cost of biases the estimated extreme precipitation changes'.

Response:

Our objective is to provide the best possible (i.e., with low uncertainty and low bias) estimates of changes in local annual extreme precipitation. Achieving this goal at the local scale requires us to use as much information about internal climate variability while avoiding biases in estimates of local extreme precipitation change. We reassert, as we have argued previously, that appropriate spatial aggregation is a sound and scientifically defensible approach, as is also affirmed by Reviewer 2. To maintain the integrity of the developed adaptative constraint method, we believe it is necessary to present the combined uncertainty reduction achieved by temperature constraints and data aggregation in the abstract.

The statement on page 17 that you mention alludes to the possibility that biases could potentially compromise estimates of local extreme precipitation changes if spatial aggregation is performed over too large an area. Reviewer 2 has suggested that we determine whether the local Rx1day distribution is distinguishable from the Rx1day distribution estimated from the spatially aggregated sample. Given that we are interested in the projected change in the long-term mean of local Rx1day, we have evaluated whether spatial data aggregation introduces discernable biases into the projected change estimates. We have therefore included a figure in the supporting information (Fig. S2, also provided below) to show that aggregating data from nine 1^ox1^o grid boxes does not discernibly alter the multi-model ensemble mean estimates of extreme precipitation changes at the vast majority of individual grid cells according to the paired t-test when conducted at the 5% significance level. These tests show that with the possible exception of grid points affected by changes in the ITCZ, rejection of the null hypothesis of no discernible bias occurs essentially randomly in space, and that it does so at rates only slightly higher than the nominal significance level of 5%, which is the expected false rejection rate when there is no bias. Consistent results are obtained when the paired Wilcoxon rank-sum test is performed to evaluate differences in the multi-model ensemble median estimates. While Fig. S2 illustrates the SSP5-8.5 emissions scenario, these findings also hold for the SSP2-4.5 and SSP1-2.6 scenarios, as shown in Figures R3-1 and R3-2, respectively.

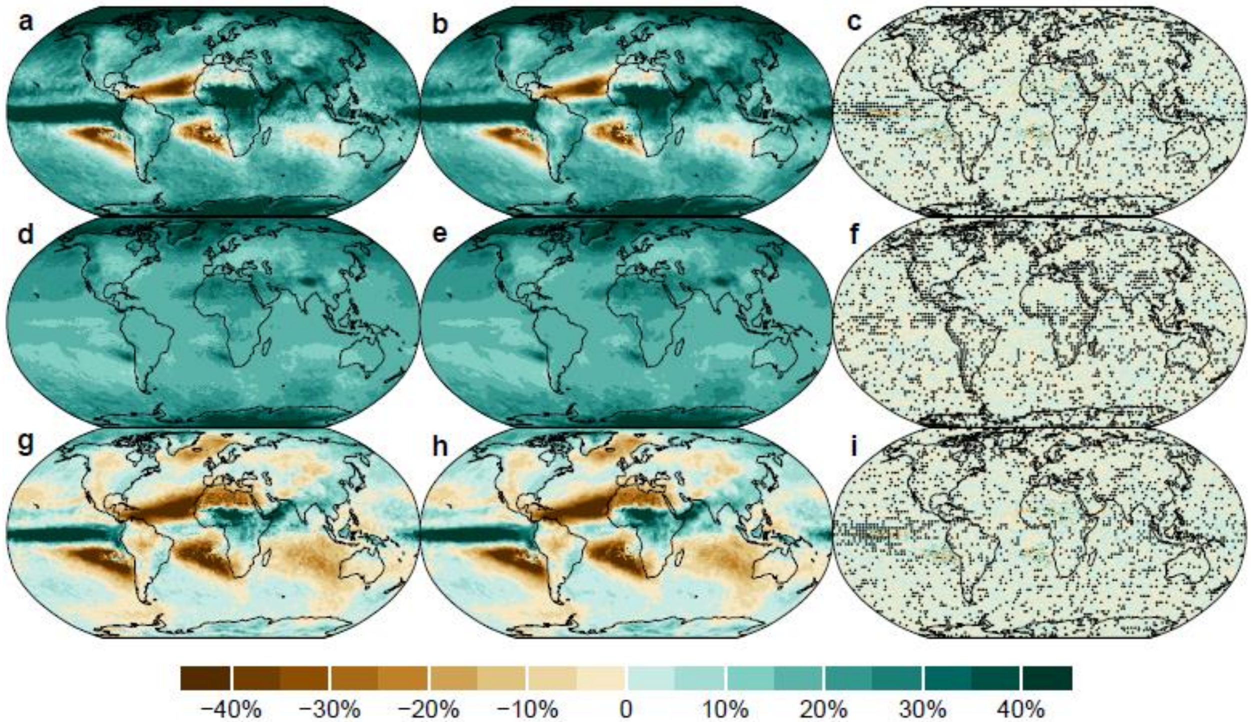

Fig. S2. The effects of spatial data aggregation on estimates of future extreme precipitation changes. The multi-model ensemble medians of the diagnosed total (a-b), thermodynamic (**d**-**e**), and dynamic changes (**g**-**h**) in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 emissions scenarios, derived with ensemble data aggregation alone (**a**, **d**, **g**) and with both ensemble and spatial data aggregation (**b**, **e**, **h**). The effects of spatial aggregation on local extreme precipitation change estimates are shown as the differences in these multi-model ensemble medians (**c**, **f**, and **i**). Stippling marks areas where the changes derived from the two data aggregation strategies are significantly different according to the paired t-test conducted at the 5% level. The fraction of such grid cells is 6.5% for total changes, 7.9% for thermodynamic changes, and 6.1% for dynamic changes.

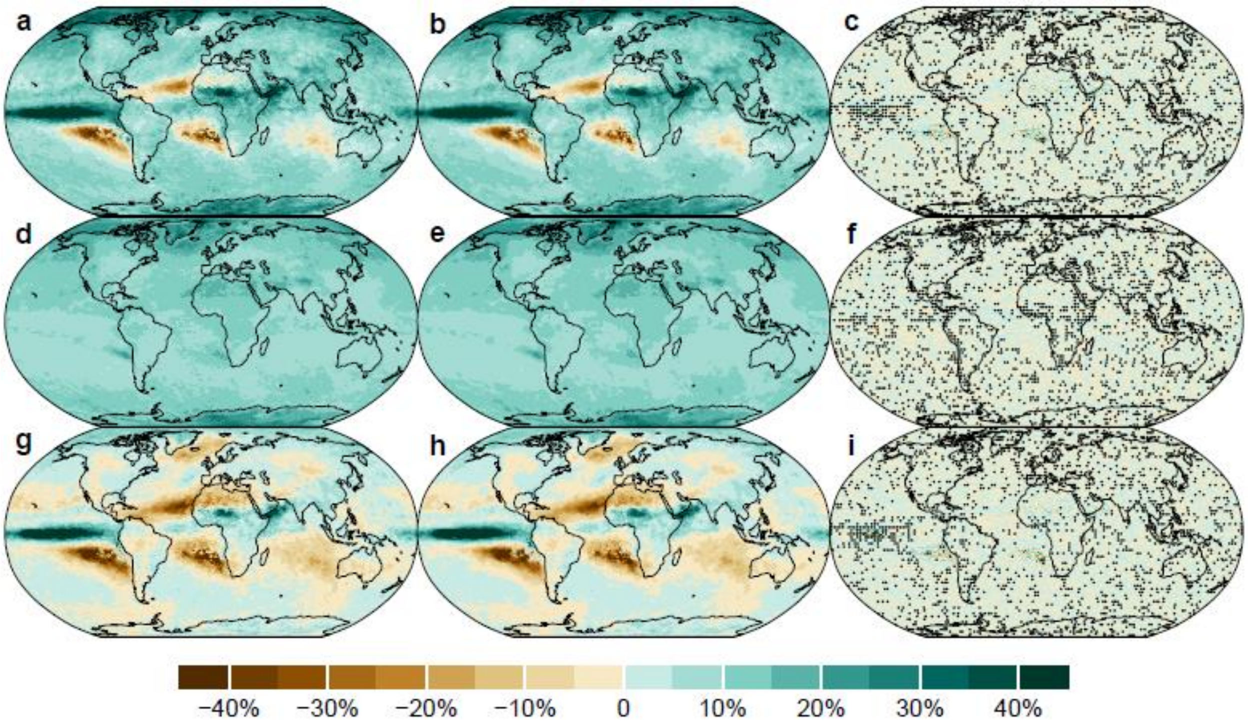

Figure R3-1. As shown in Fig. S2, but for the moderate SSP2-4.5 emissions scenario. The fraction of grid cells where spatial aggregation introduces statistically significant biases into the projected change estimates at the 5% level is 5.6% for total changes, 6.2% for thermodynamic changes, and 5.5% for dynamic changes.

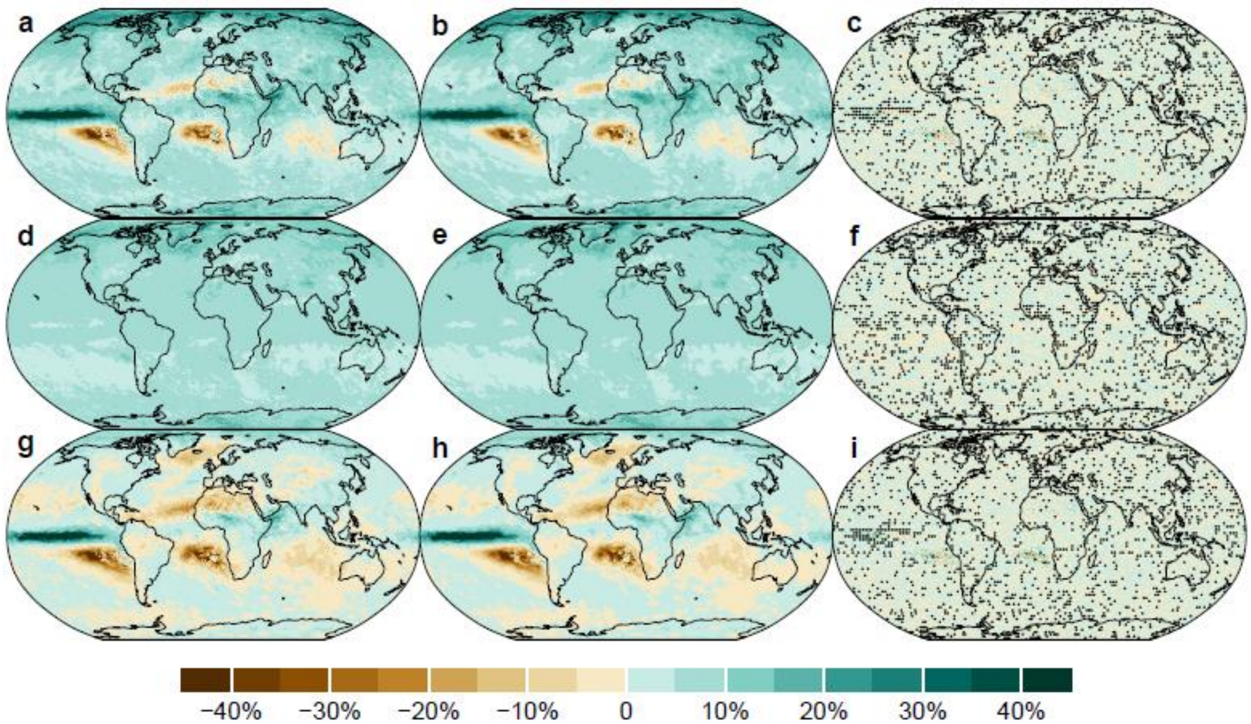

Figure R3-2. As shown in Fig. S2, but for the low SSP1-2.6 emissions scenario. The fraction of grid cells where spatial aggregation introduces statistically significant biases into the projected change estimates at the 5% level is 4.7% for total changes, 5.3% for thermodynamic changes, and 4.6% for dynamic changes.

We also include below Figures R3-3 to R3-5, which illustrate the effects on the reduction of internal variability uncertainty of ensemble aggregation alone, spatial aggregation alone, and the combined effects of ensemble and spatial aggregation.

In addition, we include Figure R3-6, which shows the combined effect of the generally very small biases documented in Figures S2, R3-1 and R3-2 and the uncertainty reductions achieved through spatial data aggregation. This is illustrated by displaying the ratio between the mean squared error (MSE) of local projections using spatial and ensemble data aggregation (in the numerator) and the MSE (i.e., variance) of local projections using ensemble data but not spatial data aggregation (in the denominator). As is evident from these figures and Figures S2, R3-1, and R3-2, spatial data aggregation makes a substantial and justified contribution to projection uncertainty reduction while inducing only modest amounts of bias in limited areas.

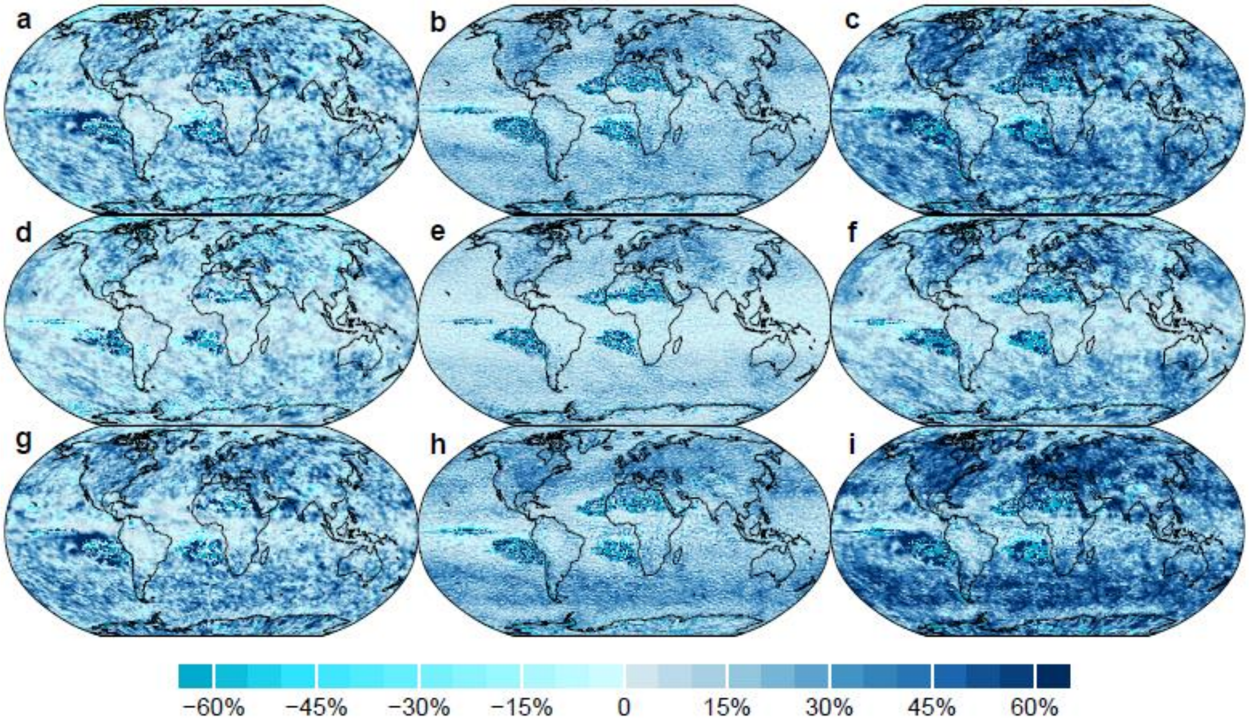

Figure R3-3. Internal variability uncertainty reductions achieved by different data aggregation strategies. The percentage reductions in the inter-model variances of the diagnosed total (first row), thermodynamic (second row), and dynamic changes (third row) in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the SSP5-8.5 emissions scenario, obtained by ensemble aggregation alone (first column), spatial aggregation alone (second column), and the combined ensemble and spatial data aggregation (third column).

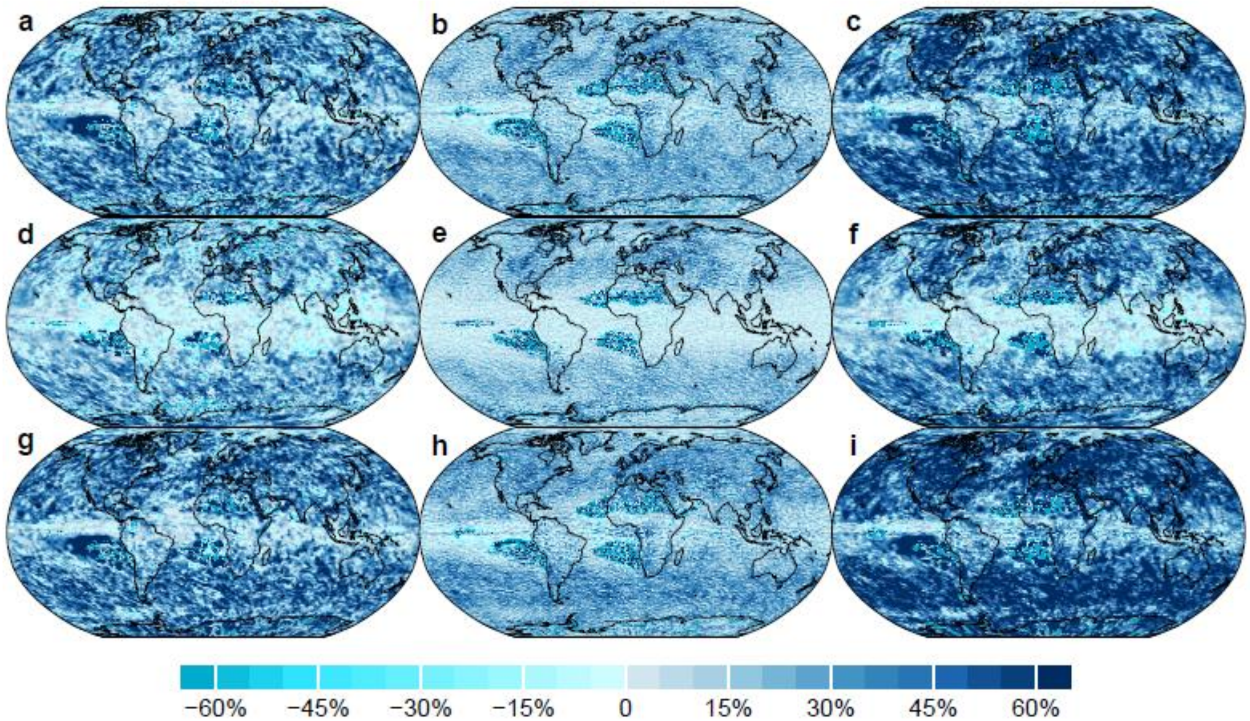

Figure R3-4. As shown in Figure R3-3, but for the moderate SSP2-4.5 emissions scenario**.**

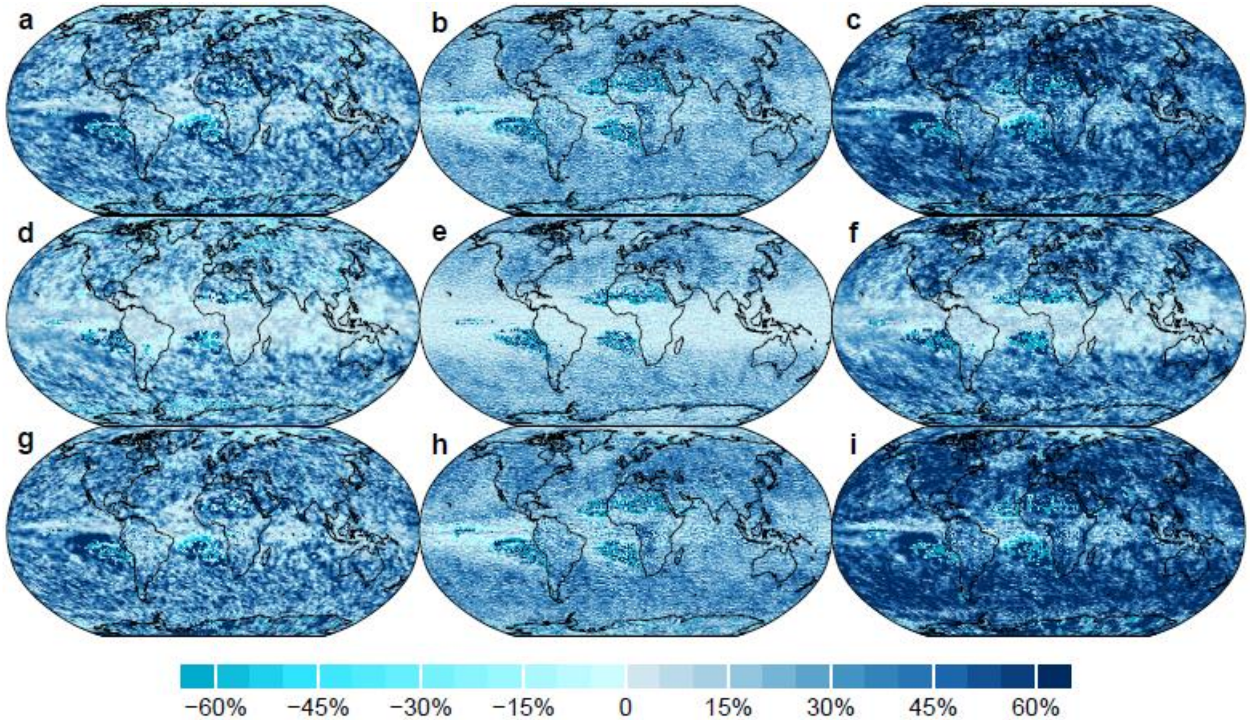

Figure R3-5. As shown in Figure R3-3, but for the low SSP2-4.5 emissions scenario**.**

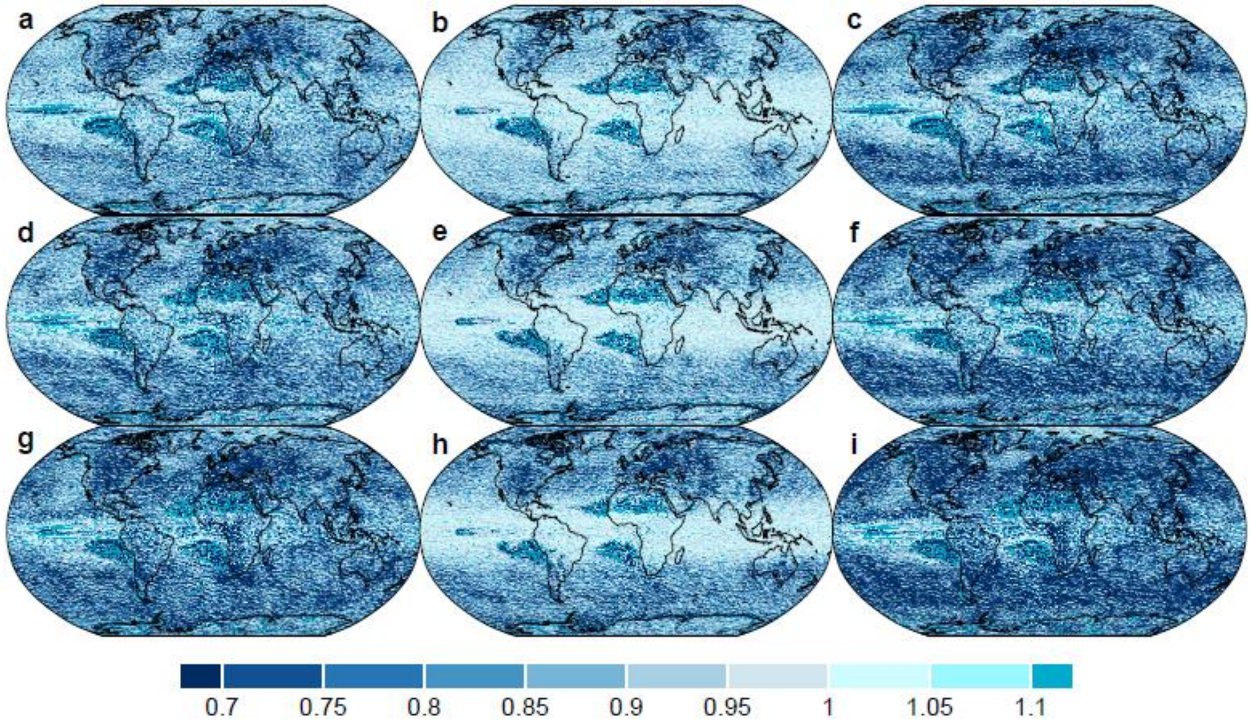

Figure R3-6. The combined effects of spatial data aggregation on estimates of future extreme precipitation changes and uncertainties. The ratios of the mean squared error of local projections of total (first column), thermodynamic (second column), and dynamic (third column) changes in annual maximum daily precipitation during 2071-2100 relative to 1985-2014 under the high SSP5-8.5 (first row), moderate SSP2-4.5 (second row), and low SSP1-2.6 (third row) emissions scenarios. These ratios compare projections obtained using ensemble and spatial data aggregation to those with ensemble aggregation only.

Reviewer #2 (Remarks to the Author):

I appreciate the authors' additional efforts to amend the manuscript. I am pleased to recommend this paper as a publication candidate for Nature Communications.

To respond to the Editor's request, I address my opinion on the method used in this study, which is the dispute between Reviewer #1 & the authors.

Reviewer #1 raised the question about the validity of results from aggregation. The average results of 3x3 adjacent grids make the percentage changes in each grid smooth, and I understand this smoothing reduces spatial uncertainty related to extreme rainfall (e.g., synoptic scale weather system).

Event rareness for extreme rainfall in limited periods (even several decades) makes sample representativity difficult. If the adjacent grids have similar characteristics to the centered grid, data aggregation is a widely used method, even in an hourly extreme rainfall study (e.g., Park & Min, 2017; Lee et al., 2022), to prepare proper numbers of samples to apply statistical analysis.

Climatologically and geologically, I expect adjacent 3x3 are likely to have a similar distribution in extreme rainfall characteristics in historical periods, which can be the climatological prerequisite for the data aggregation area threshold.

Therefore, to support the validity of data aggregation in 3x3 grids, authors may display the statistical test results (e.g., KS test) comparing the annual maximum rainfall distribution in 30 years from samples of 3x3 (in total 270 samples) and samples of the centered grid point (in total, 30 samples).

Authors can count how many models have significant differences in any grid point. Then, authors can suggest specific conditions for applying 3x3 aggregation (e.g., grids point where less than 10% GCMs have significant differences in KS-test results). This information can support data aggregation 3x3, which is not just data processing but a statistical method based on the presence of climatological characteristics.

Reference:

Park, I. H., & Min, S. K. (2017). Role of convective precipitation in the relationship between subdaily extreme precipitation and temperature. Journal of Climate, 30(23), 9527-9537.

Lee, D., Min, S. K., Park, I. H., Ahn, J. B., Cha, D. H., Chang, E. C., & Byun, Y. H. (2022). Enhanced role of convection in future hourly rainfall extremes over South Korea. Geophysical Research Letters, 49(22), e2022GL099727.

Response:

We thank the reviewer for recognizing our revisions, dedicating time and effort to critically evaluate the dispute between us and the first review, and for offering constructive suggestions. After careful consideration, we opted to directly compare the multi-model ensemble estimates of extreme precipitation changes obtained with and without spatial data aggregation. Using paired t-tests applied to the multi-model estimates of extreme precipitation changes derived from the two approaches at a 5% significance level, we confirmed that aggregating data from nine 1^{o}x1^{o} grid boxes does not significantly alter the multi-model mean estimates across the vast majority of individual grid cells. Across all considered emissions scenarios, the global rate of rejection of the null hypothesis of no alteration is close the expected value of 5% (see our response to Reviewer 1's Comment 2 for additional details). Additionally, we compared the mean squared error (MSE) of the estimates using ensemble and spatial data aggregation to the MSE of the estimates with ensemble data aggregation alone, further confirming that the adopted spatial data aggregation strategy makes a substantial and justified contribution to projection uncertainty reduction while inducing only modest amounts of bias in limited areas (again, please see our response to Reviewer 1's Comment 2 for more details). These results have also been incorporated into the revised manuscript.

For example:

In the Results section:

"*The aggregation of data from adjacent grid cells does not induce considerable biases in the estimates of local extreme precipitation changes (Fig. S2).*"

In the Methods sections,

> "*The 3x3 grid cell aggregation strategy does not introduce significant biases in the estimates of grid cell extreme precipitation changes at the vast majority of individual grid cells, as confirmed by paired t-tests at the 5% level (Fig. S2).*"

Reviewer Comments

Paper: "Increasing certainty in projected local extreme precipitation change", Li et al.

Prepared for: Nature Climate Change

The authors propose novel methodology for (1) diagnosing dynamic and thermodynamic changes to precipitation, (2) data aggregation of these changes to boost signal-to-noise, and (3) constraining local projections of extreme precipitation using a temperature-based emergent constraint that reduces model uncertainty. In general this manuscript is well-written and clearly describes the methods and results. Furthermore, there is clear value with respect to refining local projections of extreme precipitation, since this is the spatial scales at which adaptation takes place.

There are several issues that I would appreciate a response from the authors on.

1. Increased certainty versus "correctness" of projections from uncertain models

   As is often the case with this sort of message, I'm wary of advertising "increased certainty" as a positive. For example, in some cases reduced uncertainty would be a bad thing if the "true" uncertainty were quite large or, more seriously, if the quantity about which the uncertainty refers to is just plain wrong. I appreciate that you validate your methodology rigorously and ensure that, e.g., the nominal coverage is maintained (which I agree confirms that the MME is not overconfident, at least relative to itself). Ultimately, my concern here is that we present the scientific community with "better" projections (i.e., reduced uncertainty) but the projections are wrong. Particularly for local changes to extreme precipitation, models are famously bad at simulating precipitation changes (see, e.g., Polade et al., 2014, 2017; Gershunov et al., 2019, and many of your references) – which I realize is a motivation for the study in the first place. A recent paper even found that within the United States, many GCMs get the sign wrong for simulating changes in summertime precipitation extremes in the historical period (Risser et al., 2024). You acknowledge this point on line 255: "It should also be recognized that while cross-validation against pseudo-observations from global climate models very substantially increases confidence in the constrained projections, they remain subject to the limitations of process representation in global climate models and thus the realism with which the global models simulate extreme precipitation events."

I'm not completely sure what to recommend. One thing that I think would be helpful is an expanded discussion on this in the conclusions, making it more explicit that even with "increased certainty" in projections, the projections themselves could be quite wrong.

2. Treatment of anthropogenic aerosols and the fast precipitation response

   One reason that GCMs have such large uncertainties in simulating local precipitation change is their treatment of anthropogenic aerosols (Hegerl et al., 2015), in both their internal representation of aerosols (e.g., prescribed versus parameterized) as well as the precipitation response to aerosols. As far as I can tell, the only way your method accounts for anthropogenic aerosols is via the slow response to global aerosol forcing as represented by the decreased radiative forcing in the GSAT time series over 1981-2020. This completely ignores the fast precipitation response to aerosols, which has been shown to be significant even in the historical period in both observations (Tao et al., 2012; Risser et al., 2024) and models (see the PDRMIP papers; Samset et al., 2016; Myhre et al., 2017). There is a growing body of papers outlining how and why one must account for aerosols when trying to assess changes to regional precipitation (Hegerl et al., 2015; Sarojini et al., 2016; Persad et al., 2022). The fact that you ignore the fast precipitation response to aerosols feels like a major limitation for this study, and one that is furthermore not acknowledged anywhere in the manuscript.

3. Annual perspective vs. importance of seasonality for extreme precipitation

   I understand why you take an annual perspective on projected changes to local extreme precipitation, but in my view this completely ignores the important seasonal properties of extreme precipitation and, more directly, the ways that the different mechanisms (e.g., large-scale systems, ARs, convection, TCs, etc.) that generate extreme precipitation may be changing in light of external forcing. Importantly, the seasonality piece is critical for climate change adaptation and mitigation, one of the stated reasons for your study in the first place. Could you assess the seasonality in, e.g., your constrained projections in Figure 5?

4. Grid-cell testing / stippling

   Throughout your manuscript, there are many cases where you conduct a statistical test at each $1^\circ \times 1^\circ$ GCM grid box: e.g., Figures 3 and 4, and also (I think) in deciding where the adaptive constraint applies in lines 167ff. Conducting a large number of spatially-correlated hypothesis tests as in, e.g., Figure 3 is subject to the usual issues associated with multiplicity in testing (Benjamini and Hochberg, 1995; Wilks, 2016). There are simple ways to address this – see, e.g., Wilks (2016).

Otherwise, just a few minor concerns:

- (Line 194) In what sense is the RMSE a combination of bias and uncertainty? Squared errors are just an assessment of the bias – this is evident in Figure 4, where the "bias reduction" and "RMSE reduction" have nearly identical patterns.
- Also regarding Figure 4: a small issue with the terminology, although this may be a personal bias. The y-axis in each subpanel is labeled "reduction" – I think "change" might be better? As is, you have a negative reduction, which could imply that the bias, uncertainty, etc. is actually increased.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Gershunov, A., Shulgina, T., Clemesha, R. E., Guirguis, K., Pierce, D. W., Dettinger, M. D., Lavers, D. A., Cayan, D. R., Polade, S. D., Kalansky, J., et al. (2019). Precipitation regime change in western north america: the role of atmospheric rivers. *Scientific reports*, 9(1):1–11.

Hegerl, G. C., Black, E., Allan, R. P., Ingram, W. J., Polson, D., Trenberth, K. E., Chadwick, R. S., Arkin, P. A., Sarojini, B. B., Becker, A., et al. (2015). Challenges in quantifying changes in the global water cycle. *Bulletin of the American Meteorological Society*, 96(7):1097–1115.

Myhre, G., Forster, P. M., Samset, B. H., Hodnebrog, Ø., Sillmann, J., Aalbergsjø, S. G., Andrews, T., Boucher, O., Faluvegi, G., Flaschner, D., Iversen, T., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J.-F., Olivié, D., Richardson, T. B., Shindell, D., Shine, K. P., Stjern, C. W., Takemura, T., Voulgarakis, A., and Zwiers, F. (2017). PDRMIP: A precipitation driver and response model intercomparison project—protocol and preliminary results. *Bulletin of the American Meteorological Society*, 98(6):1185 – 1198.

Persad, G. G., Samset, B. H., and Wilcox, L. J. (2022). Aerosols must be included in climate risk assessments. *Nature*, 611(7937):662–664.

Polade, S. D., Gershunov, A., Cayan, D. R., Dettinger, M. D., and Pierce, D. W. (2017). Precipitation in a warming world: Assessing projected hydro-climate changes in California and other Mediterranean climate regions. *Scientific reports*, 7(1):1–10.

Polade, S. D., Pierce, D. W., Cayan, D. R., Gershunov, A., and Dettinger, M. D. (2014). The key role of dry days in changing regional climate and precipitation regimes. *Scientific reports*, 4(1):1–8.

Risser, M. D., Collins, W. D., Wehner, M. F., O'Brien, T. A., Huang, H., and Ullrich, P. A. (2024). Anthropogenic aerosols mask increases in US rainfall by greenhouse gases. *Nature Communications*.

Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Ø., Andrews, T., Faluvegi, G., Fläschner, D., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J.-F., Olivié, D., Richardson, T., Shindell, D., Shine, K. P., Takemura, T., and Voulgarakis, A. (2016). Fast and slow precipitation responses to individual climate forcers: A PDRMIP multimodel study. *Geophysical Research Letters*, 43(6):2782–2791.

Sarojini, B. B., Stott, P. A., and Black, E. (2016). Detection and attribution of human influence on regional precipitation. *Nature Climate Change*, 6(7):669–675.

Tao, W.-K., Chen, J.-P., Li, Z., Wang, C., and Zhang, C. (2012). Impact of aerosols on convective clouds and precipitation. *Reviews of Geophysics*, 50(2).

Wilks, D. (2016). "the stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, 97(12):2263–2273.